

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
8 November 2001 (08.11.2001)

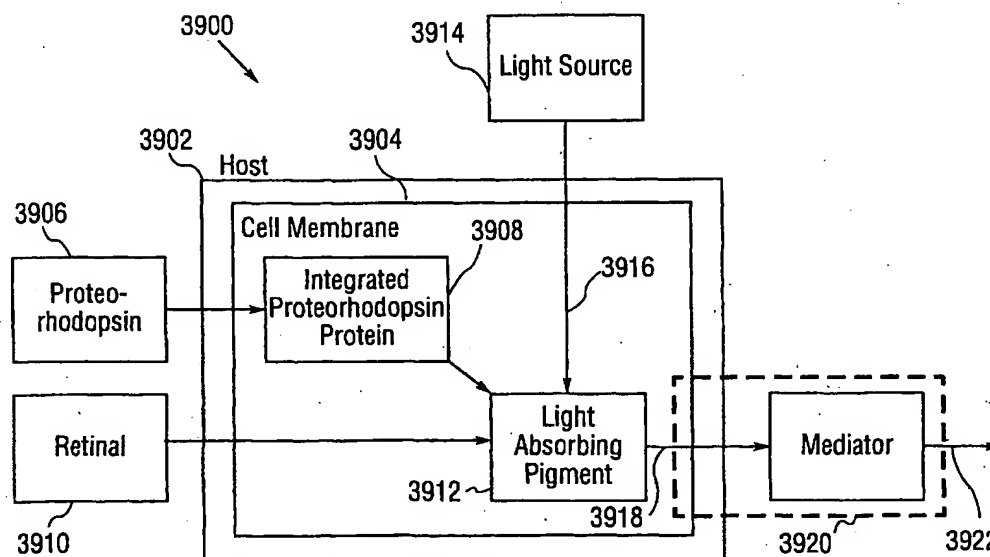
PCT

(10) International Publication Number
WO 01/83701 A2

- (51) International Patent Classification⁷: C12N (74) Agent: KAMINSKY, Rena; Lumen Intellectual Property Services, Inc., 45 Cabot Avenue, Suite 110, Santa Clara, CA 95051 (US).
- (21) International Application Number: PCT/US01/14394
- (22) International Filing Date: 2 May 2001 (02.05.2001) (81) Designated States (*national*): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CO, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW.
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 60/201,602 3 May 2000 (03.05.2000) US (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- (71) Applicant: MONTEREY BAY AQUARIUM RESEARCH INSTITUTE [US/US]; 7700 Sandholt Road, Moss Landing, CA 95039 (US).
- (72) Inventors: DELONG, Edward, F.; 4 Blacktail Lane, Monterey, CA 93940 (US). BEJA, Oded; 408 Holovits Court, Marina, CA 93933 (US).
- Published:
— without international search report and to be republished upon receipt of that report

[Continued on next page]

(54) Title: LIGHT-DRIVEN ENERGY GENERATION USING PROTEORHODOPSIN



(57) Abstract: A light-driven energy generation system using proteorhodopsin is provided. Proteorhodopsin sequences were retrieved and amplified from naturally occurring members of the domain Bacteria using proteorhodopsin-specific polymerase chain reaction primers. Proteorhodopsin sequences were placed in expression vectors for production of proteorhodopsin proteins in a host, for instance, *E. coli* and other bacteria. The system also includes a light source and a source of retinal, that allows the system to convert light into biochemical energy. The generated biochemical energy could be mediated into electrical energy by a mediator.



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

PATENT APPLICATION

LIGHT-DRIVEN ENERGY GENERATION USING PROTEORHODOPSIN

INVENTORS

Edward F. DeLong and Oded Beja

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is cross-referenced to and claims priority from U.S. Provisional application 60/201,602 filed 05/03/2000, which is hereby incorporated by reference.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

This invention was supported in part by grant number OCE 0001619 from the National Science Foundation (NSF). The U.S. government has certain rights in the invention.

STATEMENT TO COMPUTER DISK AND SEQUENCE LISTING

This application includes a sequence listing of 65 sequences and a computer disk labeled "Sequence Listing for application entitled "Light-driven energy generation using proteorhodopsin" by Edward F. DeLong and Oded Beja" containing files "MBA101-SEQLIST.prj", dated "04/23/01" with 174,089 bytes, which is the PatentIn

project file generated using PatentIn Version 3.0 software provided by the USPTO, and "MBA101-SEQLIST.txt", dated "04/23/01" with 323,739 bytes, which is the generated sequence listing from the PatentIn project file MBA101-SEQLIST.prj using PatentIn Version 3.0 software, all which are herein incorporated. The information recorded in computer readable format on the incorporated computer disk labeled "Sequence Listing" containing files "MBA101-SEQLIST.prj" and "MBA101-SEQLIST.txt" are identical to the incorporated written sequence listing.

FIELD OF THE INVENTION

The present invention relates generally to gene expression of functional recombinant proteins in bacteria. More particularly, the present invention relates to proteorhodopsin genes and proteins that function as a light-driven energy generator in *Escherichia coli* (E. coli) and other bacteria.

BACKGROUND ART

Retinal (vitamin A aldehyde) is a chromophore that binds integral membrane proteins (opsins) to form light-absorbing pigments called rhodopsins. Rhodopsins are currently known to belong to two distinct protein families. The visual rhodopsins, found in the eye throughout the animal kingdom, are photosensory pigments. Archeal rhodopsins, found in extreme halophilic environments, function as light-driven protons pumps (bacteriorhodopsins), chloride ion pumps (halorhodopsins), or photosensory receptors (sensory rhodopsins). The two protein families show no significant sequence similarity and may have different origins. They do, however, share identical topologies characterized by seven transmembrane α -helices that form a pocket in which retinal is covalently linked, as a protonated Schiff base (helix G).

The archaeal rhodopsins are able to generate a photocycle which produces a chemiosmotic membrane potential in response to light, as such light energy is converted into biochemical energy. Recently, a protein with high sequence similarity to the archaeal rhodopsins has also been retrieved in the eukaryote *Neurospora crassa* (J.A. Bieszke et al., *Proceedings of National Academy of Sciences USA* 96:8034, 1999). The eucaryal rhodopsin formed a photochemically reactive pigment when bound to all-trans retinal and exhibited photocycle kinetics similar to those of archaeal sensory rhodopsins (J.A. Bieszke et al., *Biochemistry* 38:14138, 1999). To date, however, no rhodopsin-like sequences have been reported in members of the domain Bacteria, and no light-driven proton pumps based on rhodopsin have ever before been functionally expressed in *E. coli*.

The phototropic conversion of light energy into biochemical energy using bacteriorhodopsin can be harnessed for a variety of processes and applications, such as bio-electronic applications and bio-materials, as has been reported in US Patent No. 5,757,525 for optical devices, US Patent No. 5,854,710 for optical Fourier processing, and US Patent No. 5,470,690 for optical information storage. Bacteriorhodopsin in bio-electronic applications is aimed to interface, integrate, or substitute the silicon based microelectronics systems as well as molecular devices. Bacteriorhodopsin as a bio-material is integrated, for instance, in optical films for light mediated computer memory applications and pattern recognition.

Previously, archaeal rhodopsins capable of generating a chemiosmotic membrane potential in response to light had only been found in halophilic archaea. Therefore, rhodopsins that originate from archaea adapted to highly saline environments cannot be functionally expressed in *E. coli*. Finally, the isolation and cultivation of

halorhodopsins is an elaborate process. At present one does not foresee an economic utilization possible for this process (e.g. US Patent 5,290,699).

Accordingly, as one skilled in the art might readily acknowledge, there is a strong need to retrieve and provide rhodopsin-like sequences from naturally occurring members of the domain Bacteria.

OBJECTS AND ADVANTAGES

In light of the above, it is the primary objective of the present invention to provide rhodopsin-like sequences from naturally occurring members of the domain Bacteria. More specifically, it is the objective of the present invention to provide a method to retrieve proteorhodopsin genes from DNA of naturally occurring bacteria that encodes DNA sequence for proteorhodopsin proteins.

It is another objective of the present invention to provide proteorhodopsin-specific polymerase chain reaction primers that amplify the proteorhodopsin-containing gene from a DNA sample of naturally occurring bacteria.

It is yet another objective of the present invention to produce variants of a proteorhodopsin gene using the same proteorhodopsin-specific polymerase chain reaction primers by amplifying a proteorhodopsin-containing gene from of a mixed sample of naturally occurring bacteria.

It is still another objective of the present invention to provide an expression vector that produces a proteorhodopsin protein in *E. coli* and other bacteria.

It is another objective of the present invention to provide a light-driven energy generator in which the functional properties of proteorhodopsin are utilized. These properties include the ability to integrate within a host, for instance a cell membrane of *E. coli*, making an integrated proteorhodopsin protein, and the ability to bind retinal, making a light absorbing pigment.

It is another objective of the present invention to provide a light source and illuminate the light absorbing pigment to convert light energy into biochemical energy.

It is another objective of the present invention to provide a mediator and mediate the biochemical energy into electrical energy.

It is another objective of the present invention to provide methods to manipulate the kinetics of the light-driven energy generator.

The advantage of the present invention over the prior art is that it is not restricted to operate in halophilic archaea and could therefore be functionally expressed in *E. coli* and other bacteria. Accordingly, another advantage of the present invention is that it provides for a fast and cheap production method that allows for mass production of functionally active proteorhodopsin.

SUMMARY

The present invention provides proteorhodopsin gene and protein sequences retrieved from samples of naturally occurring members of the domain Bacteria. More specifically, the present invention provides a method for the retrieval and amplification of proteorhodopsin genes from DNA samples of naturally occurring marine bacteria. In accordance with several exemplary embodiments of the present invention, DNA samples were obtained from naturally occurring bacteria such as, for instance, marine proteobacteria, SAR86 bacteria, or recombinant DNA libraries containing naturally occurring bacteria. The present invention provides proteorhodopsin-specific polymerase chain reaction (PCR) primers to amplify a proteorhodopsin gene from DNA samples of these marine bacteria. The present invention also provides a device and method for the placement of proteorhodopsin genes in an expression vector to produce functional proteorhodopsin proteins in *E. coli* and other bacteria.

Accordingly, the present invention provides a method to produce and obtain variants of proteorhodopsin genes and proteins. The same proteorhodopsin-specific polymerase chain reaction primers amplify different variants of proteorhodopsin-containing genes from a mixed sample of naturally occurring bacteria. As one skilled in the art might readily acknowledge, these variants of a proteorhodopsin gene produce functional variations in the photocycle kinetics of the proteorhodopsin protein.

Furthermore, the present invention provides a light-driven energy generator that utilizes proteorhodopsin to convert light-energy into biochemical energy. This light-driven energy generator takes advantage of the functional properties of the proteorhodopsin protein once expressed in, for example, *E. coli* or other bacteria as is

described in exemplary embodiments. These properties include the ability to integrate within a host such as, for instance, a cell membrane of *E. coli* or other Bacteria, and thereby making an integrated proteorhodopsin protein or integrated cell membrane protein. These properties also include the ability to bind retinal and thereby making a light absorbing pigment. Illuminating the light absorbing pigment with a light source converts light energy into biochemical energy. Finally, the biochemical energy can be mediated into electrical energy by a mediator.

In accordance with exemplary embodiments, the present invention enables one skilled in the art to manipulate the kinetics of the proteorhodopsin protein photocycle once it is operational in the light-driven energy generator. In particular, the present invention provides examples in which the light source characteristics are manipulated. Examples are the manipulation of the delivery of fast-light pulses and/or the delivery of light at different wavelengths. The present invention also provides examples in which incremental additions of retinal influences the function of the light-driven energy generator. In addition, a proteorhodopsin gene or protein variant can be selected to determine an absorption spectra of the light absorbing pigment to change the kinetics of the light energy generator, for instance to meet a design/functional criteria of an application wherein proteorhodopsin is utilized.

BRIEF DESCRIPTION OF THE FIGURES

The objectives and advantages of the present invention will be understood by reading the following detailed description in conjunction with the drawings, in which:

- FIG. 1** illustrates the phylogenetic tree of bacterial 16S rRNA gene sequences including that encoded on the 130 kb bacterioplankton BAC clone (EBAC31A8).
- FIG. 2** provides a nucleotide sequence of polymerase chain reaction primer 1 (Sequence ID No:2) used to amplify a proteorhodopsin gene.
- FIG. 3** provides a nucleotide sequence of polymerase chain reaction primer 2 (Sequence ID No:3) used to amplify a proteorhodopsin gene.
- FIG. 4** provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:4) amplified from clone EBAC31A8 (Sequence ID No:1) using PCR primers 1 (Sequence ID No:2) and 2 (Sequence ID No:3), and the deduced amino acid sequence (Sequence ID No:5) of the proteorhodopsin gene Sequence ID No:4 amplified from clone EBAC31A8 (Sequence ID No:1).
- FIG. 5** provides a map of the secondary structure of the proteorhodopsin protein (Sequence ID No:7). Single letter amino acid codes are used (according to J. Sasaki and J.L. Spudich, Biophys. J. 75:2435, 1998). Predicted retinal binding pocket residues are marked in black.
- FIG. 6** provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:8) amplified from clone EBAC40E8 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:9) of the proteorhodopsin gene Sequence ID No:8 amplified from clone EBAC40E8.
- FIG. 7** provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:10) amplified from clone EBAC41B4 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:11) of the proteorhodopsin gene Sequence ID No:7 amplified from clone EBAC41B4.

FIG. 8 provides the nucleotide sequence of the proteorhodopsin gene (**Sequence ID No:12**) amplified from clone EBAC64A5 using PCR primers 1 (**Sequence ID No:2**) and 2 (**Sequence No:3**), and the deduced amino acid sequence (**Sequence ID No:13**) of the proteorhodopsin gene **Sequence ID No:12** amplified from clone EBAC64A5.

FIG. 9 provides a variants map of the DNA sequences of the proteorhodopsin gene with **Sequence ID No:4**, **Sequence ID No:8**, **Sequence ID No:10**, and **Sequence ID No:12** that were amplified from clone EBAC38A8, EBAC40E8, EBAC41B4 and EBAC64A5 respectively using the proteorhodopsin-specific PCR primer 1 (**Sequence ID No:2**) and 2 (**Sequence ID No:3**). Dots represent sequences having identical sequence as those in **Sequence ID No:4**.

FIG. 10 provides a variant map of the deduced amino acid sequences encoded by the proteorhodopsin gene with **Sequence ID No:4**, **Sequence ID No:8**, **Sequence ID No:10**, and **Sequence ID No:12** that were amplified from respectively EBAC38A8, EBAC40E8, EBAC41B4 and EBAC64A5 using the proteorhodopsin-specific primer 1 (**Sequence ID No:2**) and 2 (**Sequence ID No:3**). Lower case represents the PCR primer sequence region. Dots represent residues having identical sequence as those in **Sequence ID No:5**.

FIG. 11 provides the nucleotide sequence of the proteorhodopsin gene (**Sequence ID No:14**) amplified from clone HOT0m1 using PCR primers 1 (**Sequence ID No:2**) and 2 (**Sequence No:3**), and the deduced amino acid sequence (**Sequence ID No:15**) of the proteorhodopsin gene **Sequence ID No:14** amplified from clone HOT0m1.

FIG. 12 provides the nucleotide sequence of the proteorhodopsin gene (**Sequence ID No:16**) amplified from clone HOT75m1 using PCR primers 1 (**Sequence ID**

No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:17) of the proteorhodopsin gene Sequence ID No:16 amplified from clone HOT75m1.

FIG. 13 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:18) amplified from clone HOT75m3 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:19) of the proteorhodopsin gene Sequence ID No:18 amplified from clone HOT75m3.

FIG. 14 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:20) amplified from clone HOT75m4 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:21) of the proteorhodopsin gene Sequence ID No:20 amplified from clone HOT75m4.

FIG. 15 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:22) amplified from clone HOT75m8 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:23) of the proteorhodopsin gene Sequence ID No:22 amplified from clone HOT75m8.

FIG. 16 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:24) amplified from clone MB0m1 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:25) of the proteorhodopsin gene Sequence ID No:24 amplified from clone MB0m1.

FIG. 17 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:26) amplified from clone MB0m2 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence

(Sequence ID No:27) of the proteorhodopsin gene Sequence ID No:26 amplified from clone MB0m2.

FIG. 18 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:28) amplified from clone MB20m2 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:29) of the proteorhodopsin gene Sequence ID No:28 amplified from clone MB20m2.

FIG. 19 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:30) amplified from clone MB20m5 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:31) of the proteorhodopsin gene Sequence ID No:30 amplified from clone MB20m5.

FIG. 20 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:32) amplified from clone MB20m12 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:33) of the proteorhodopsin gene Sequence ID No:32 amplified from clone MB20m12.

FIG. 21 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:34) amplified from clone MB40m1 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:35) of the proteorhodopsin gene Sequence ID No:34 amplified from clone MB40m1.

FIG. 22 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:36) amplified from clone MB40m5 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence

(Sequence ID No:37) of the proteorhodopsin gene Sequence ID No:36 amplified from clone MB40m5.

FIG. 23 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:38) amplified from clone MB40m12 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:39) of the proteorhodopsin gene Sequence ID No:38 amplified from clone MB40m12.

FIG. 24 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:40) amplified from clone MB100m5 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:41) of the proteorhodopsin gene Sequence ID No:40 amplified from clone MB100m5.

FIG. 25 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:42) amplified from clone MB100m7 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:43) of the proteorhodopsin gene Sequence ID No:42 amplified from clone MB100m7.

FIG. 26 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:44) amplified from clone MB100m9 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:45) of the proteorhodopsin gene Sequence ID No:44 amplified from clone MB100m9.

FIG. 27 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:46) amplified from clone MB100m10 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence

(Sequence ID No:47) of the proteorhodopsin gene Sequence ID No:46 amplified from clone MB100m10.

FIG. 28 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:48) amplified from clone PALB1 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:49) of the proteorhodopsin gene Sequence ID No:48 amplified from clone PALB1.

FIG. 29 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:50) amplified from clone PALB2 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:51) of the proteorhodopsin gene Sequence ID No:50 amplified from clone PALB2.

FIG. 30 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:52) amplified from clone PALB5 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:53) of the proteorhodopsin gene Sequence ID No:52 amplified from clone PALB5.

FIG. 31 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:54) amplified from clone PALB7 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:55) of the proteorhodopsin gene Sequence ID No:54 amplified from clone PALB7.

FIG. 32 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:56) amplified from clone PALB6 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence

(Sequence ID No:57) of the proteorhodopsin gene Sequence ID No:56 amplified from clone PALB6.

FIG. 33 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:58) amplified from clone PALB8 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:59) of the proteorhodopsin gene Sequence ID No:58 amplified from clone PALB8.

FIG. 34 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:60) amplified from clone PALE1 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:61) of the proteorhodopsin gene Sequence ID No:60 amplified from clone PALE1.

FIG. 35 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:62) amplified from clone PALE6 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:63) of the proteorhodopsin gene Sequence ID No:62 amplified from clone PALE6.

FIG. 36 provides the nucleotide sequence of the proteorhodopsin gene (Sequence ID No:64) amplified from clone PALE7 using PCR primers 1 (Sequence ID No:2) and 2 (Sequence No:3), and the deduced amino acid sequence (Sequence ID No:65) of the proteorhodopsin gene Sequence ID No:64 amplified from PALE7.

FIG. 37 illustrates a phylogenetic tree of different proteorhodopsin genes.

FIG. 38 provides an example of an alignment of proteorhodopsin amino acid sequences.

FIG. 39 provides a light-driven energy generator that utilizes proteorhodopsin.

FIG. 40 provides an example of a proteorhodopsin-expressing *E. coli* cell suspension (+) compared to control cells (-), both with all-*trans* retinal.

FIG. 41 provides an example of absorption spectra of retinal-constituted proteorhodopsin in *E. coli* membranes and a negative control.

FIG. 42 provides an example of a light-driven transport of protons by a proteorhodopsin-expressing *E. coli* cell suspension.

FIG. 43 provides an example of a transport of [³H]TPP⁺ in *E. coli* right-side-out vesicles containing expressed proteorhodopsin, reconstituted with or without 10 μM retinal in the presence of light or in the dark.

FIG. 44 provides an example of laser flash-induced absorbance changes in suspensions of *E. coli* membranes containing proteorhodopsin.

FIG. 45 provides an example of absorption spectra of retinal-constituted proteorhodopsin in *E. coli* membranes.

DETAILED DESCRIPTION

Although the following detailed description contains many specifics for the purposes of illustration, anyone of ordinary skill in the art will appreciate that many variations and alterations to the following details are within the scope of the invention. Accordingly, the following preferred embodiment of the invention is set forth without any loss of generality to, and without imposing limitations upon, the claimed invention.

Proteorhodopsin

The present invention provides rhodopsin-like gene and protein sequences retrieved from naturally occurring members of the domain Bacteria. More specifically, the

present invention provides a method for the retrieval and amplification of proteorhodopsin genes from DNA samples of naturally occurring marine bacteria. In accordance with exemplary embodiments of the present invention, DNA samples were obtained from naturally occurring marine bacteria such as bacteria from the SAR86 group. Provided as an exemplary embodiment of the SAR86 group, DNA samples were obtained from a bacterioplankton Bacterial Artificial Chromosome (BAC) clone BAC31A8 (also referred to as EBAC31A08). In general, as will be appreciated by those of ordinary skill in the art, suitable DNA samples can also be obtained from other sources, e.g., from a marine environment or from a recombinant DNA library containing genomic fragments of samples of naturally occurring bacteria.

FIG. 1 shows the phylogenetic tree of bacterial 16S rRNA gene sequences including that encoded on the EBAC31A8. **FIG. 1** also shows the relationship of EBAC31A8 to the SAR86 bacteria group as well as to the gamma-proteobacteria group. A subclone shotgun library was constructed from BAC clone 31A8, and subclones were sequenced in both directions on the MegaBACE 1000 capillary array electrophoresis DNA sequencing instrument (Molecular Dynamics, Sunnyvale, CA). Sequence analysis of a 130-kb genomic DNA that encodes the ribosomal RNA operon from BAC31A8, reveals an open reading frame encoding a proteorhodopsin. In an exemplary embodiment, the contiguous sequence was assembled using SEQUENCHER 3.1.1 software (Gene Codes Co., Ann Arbor, MI). Other sequencing techniques can also be used, as will be recognized by those skilled in the art. The sequence of the proteorhodopsin-containing contig has been deposited in GenBank under accession #AF279106 and deposit date October 23rd, 2000. **Appendix A**, hereby incorporated, shows the nucleotide sequence of the BAC clone BAC31A8 (**Sequence ID No:1**)

which contains the 130 kilobases genomic DNA from a naturally occurring marine bacterium.

Proteorhodopsin was amplified from the 130 kilobase bacterioplankton BAC clone 31A8 (**Sequence ID No:1**) by polymerase chain reaction (PCR), using the proteorhodopsin-specific primers 5'-aCCATGGgtaaattattactgatattagg-3' (**Sequence ID No:2** and shown in **FIG. 2**) and 5'-agcattagaagattctttaacagc-3' (**Sequence ID No:3** and shown in **FIG. 3**). References for PCR are, for instance, *The Polymerase Chain Reaction*, Mullis et al., Ed. (Birkhauser, Boston, 1994) and U.S. Patent Nos. 4,683,195 and 4,683,202 to Mullis et al. The proteorhodopsin-specific PCR primers include the addition of 3 nucleotides that encoded one amino acid not found in the native gene sequence of clone BAC31A8 (**Sequence ID No:6**), in the second amino acid position which is a glycine located on the 2nd codon ("GGT"). Therefore, compare the second amino acid position in the **Sequence ID No:5** using PCR primers 1 and 2 with the native **Sequence ID no:7**. This addition of one non-native amino acid created a new restriction endonuclease site (NcoI site) not present in the native sequence. This allowed subcloning of the amplified fragment into the NcoI restriction site of an expression vector pBAD TOPO TA Cloning[®] Kit (Invitrogen, La Jolla, CA). The present invention is not limited to the use of this type of expression vector and other expression vectors could also be used.

FIG. 4 shows the nucleotide sequence of the proteorhodopsin gene (**Sequence ID No:4**) that results from amplification of the proteorhodopsin-containing DNA in BAC31A8 using proteorhodopsin-specific PCR primers **Sequence ID No:2** and **Sequence No:3**. **FIG. 4** also shows the deduced amino acid sequences (**Sequence ID No:5**) encoded by the proteorhodopsin gene (**Sequence ID No:4**).

FIG. 5 shows an exemplary embodiment of a secondary structure of proteorhodopsin after it has been folded in a cell membrane **510** and bonded with retinal **520**. **FIG. 5** shows the native proteorhodopsin gene (**Sequence ID No:6**) obtained from clone BAC31A8 and encodes a proteorhodopsin protein of 249 amino acids with a molecular weight of 27 kD (**Sequence ID No:7**). In **FIG. 5**, **530** indicates seven transmembrane domains, a typical feature of the rhodopsin protein family, that aligned well with the corresponding helices of the archaeal rhodopsins. **FIG. 5** also shows the amino acid residues that form a retinal binding pocket indicated by **520**. Although the proteorhodopsin proteins shown in **FIGS. 4** and **5** both originate from BAC31A8, they differ with respect to the second amino acid position. The reason is that the proteorhodopsin-specific PCR primers that were used to amplify the proteorhodopsin gene from BAC31A8 (which resulted in proteorhodopsin protein as in **FIG. 4**; **Sequence ID No:5**) included the addition of 3 nucleotides. These 3 nucleotides encoded one amino acid not found in the native gene sequence (**Sequence ID No:6**), in the second amino acid position which is a glycine located on the 2nd codon ("GGT"). Proteorhodopsin protein (**Sequence ID No:7**) as shown in **FIG. 5** originates from the native gene sequence without the addition of the 3 nucleotides. As mentioned above, the addition of the 3 nucleotides created a new restriction endonuclease site (NcoI site) that was not present in the native sequence and thereby allowed the amplified fragment to be subcloned into the NcoI site of the expression vector.

In the exemplary embodiment presented above, PCR primers with **Sequence ID No:2** and **Sequence ID No:3** were used. In general, the present invention provides a method for designing different proteorhodopsin-specific PCR primers that are all capable of amplifying a proteorhodopsin gene from DNA samples of naturally occurring microbial populations by polymerase chain reaction. In designing these

primers one first needs to determine a DNA sequence of a proteorhodopsin gene. Then one can design oligodeoxynucleotide primers with a Watson-Crick base pair complementary to 5' and 3' ends of the proteorhodopsin gene.

Variants of Proteorhodopsin

In the previous section, an exemplary embodiment is provided of a proteorhodopsin gene and protein. The present invention also provides the retrieval of genetic variations of proteorhodopsin from naturally occurring genetic variations in naturally occurring bacterial populations. These genetic variations in proteorhodopsin sequences result in functional variations in the proteorhodopsin proteins as is discussed below.

The present invention enables one skilled in the art to use the same proteorhodopsin-specific PCR primers as shown in FIGS. 2 and 3 to successfully amplify different sequence variants from DNA originating from mixed naturally occurring bacterial populations when it is compared to for instance the proteorhodopsin gene as shown in FIG. 4. As mentioned above, different proteorhodopsin-specific PCR primers could be used to amplify genetic variants of proteorhodopsin.

FIGS. 6-8 show exemplary embodiments of three different and unique variants of the proteorhodopsin gene that were retrieved from a recombinant DNA library of other naturally occurring bacteria (i.e. the bacterial artificial chromosome library (BAC)). In general, genetic variants could be obtained from different DNA libraries containing naturally occurring bacteria as well as from samples of naturally occurring bacteria. FIG. 6 shows the variant of the proteorhodopsin gene sequence (Sequence ID No:8) that is amplified from the BAC clone 40 (BAC40E8) with the same proteorhodopsin-

specific PCR primers as provided in **Sequence ID No:2** and **3**. Accordingly, **FIG. 6** also shows the deduced amino acid sequence (**Sequence ID No:9**) of the genetic variant of proteorhodopsin shown in **FIG. 6**. **FIG. 7** shows the variant of the proteorhodopsin gene sequence (**Sequence ID No:10**) that is amplified from the BAC clone 41 (BAC41B4) with the same proteorhodopsin-specific PCR primers as provided in **Sequence ID No:2** and **3**. Accordingly, **FIG. 7** also shows the deduced amino acid sequence (**Sequence ID No:11**) of the genetic variant of proteorhodopsin shown in **FIG. 7**. **FIG. 8** shows the variant of the proteorhodopsin gene sequence (**Sequence ID No:12**) that is amplified from the BAC clone 64 (BAC64A5) with the same proteorhodopsin-specific PCR primers as provided in **Sequence ID No:2** and **3**. Accordingly, **FIG. 8** also shows the deduced amino acid sequence (**Sequence ID No:13**) of the genetic variant of proteorhodopsin shown in **FIG. 8**.

FIG. 9 provides a variants map of the nucleotide sequences of the proteorhodopsin gene **Sequence ID No:4**, **Sequence ID No:8**, **Sequence ID No:10**, and **Sequence ID No:12** amplified from respectively BAC31A8, BAC40E8, BAC41B4 and BAC64A5 using the proteorhodopsin-specific PCR primers **Sequence ID No:2** and **Sequence ID No:3**. In **FIG. 9** lower case letters represent the PCR primer sequence region. Dots represent residues having identical sequence as those in **Sequence ID No:4**. These proteorhodopsin gene sequences differ by as much as 31 nucleotides as is shown in **FIG. 10**. **FIG. 10** provides a variant map of the deduced amino acid sequences of the proteorhodopsin genes shown in **FIG. 9**.

Using the same proteorhodopsin-specific PCR primers, as for instance shown in **FIGS. 2** and **3**, proteorhodopsin genes were also amplified from bacterioplankton extracts. As mentioned above, any proteorhodopsin-specific PCR primer can be used. These bacterioplankton extracts include those from the Monterey Bay (referred to as MB

clones), the Southern Ocean (Palmer Station, referred to as PAL clones), and waters of the central North Pacific Ocean (Hawaii Ocean Time series station, referred to as HOT clones).

FIGS. 11-36 show exemplary embodiments of different and unique variants of proteorhodopsin that were retrieved from the MB clones, PAL clones, and HOT clones. **FIGS. 11-36** each show a variant of a proteorhodopsin gene sequence that is amplified with the same proteorhodopsin-specific PCR primers as provided in **Sequence ID No:2** and **Sequence ID No:3** from respectively clones HOT0m1, HOT75m1, HOT75m3, HOT75m4, HOT75m8, MB0m1, MB0m2, MB20m2, MB20m5, MB20m12, MB40m1, MB40m5, MB40m12, MB100m5, MB100m7, MB100m9, MB100m10, PALB1, PALB2, PALB5, PALB7, PALB6, PALB8, PALE1, PALE6 and PALE7. The proteorhodopsin gene sequences retrieved from clones HOT0m1, HOT75m1, HOT75m3, HOT75m4, HOT75m8, MB0m1, MB0m2, MB20m2, MB20m5, MB20m12, MB40m1, MB40m5, MB40m12, MB100m5, MB100m7, MB100m9, MB100m10, PALB1, PALB2, PALB5, PALB7, PALB6, PALB8, PALE1, PALE6 and PALE7, have respectively **Sequence ID Nos: 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, and 64**. Accordingly, **FIGS. 11-36** also show the deduced amino acid sequence of each genetic variant of proteorhodopsin. The deduced amino acid sequence encoded by the proteorhodopsin gene retrieved from clones HOT0m1, HOT75m1, HOT75m3, HOT75m4, HOT75m8, MB0m1, MB0m2, MB20m2, MB20m5, MB20m12, MB40m1, MB40m5, MB40m12, MB100m5, MB100m7, MB100m9, MB100m10, PALB1, PALB2, PALB5, PALB7, PALB6, PALB8, PALE1, PALE6 and PALE7, have respectively **Sequence ID Nos: 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, and 65**.

In an exemplary embodiment shown in **FIG. 37**, fifteen different variants of proteorhodopsin in the PCR generated MB gene library **3710** were detected, falling into three clusters. The MB gene library includes MB clones MB0m2, MB40m5, MB20m2, MB40m12, MB100m10, MB20m12, MB40m1, MB100m5, MB20m5, MB100m7, MB0m1, and MB100m9 as well as BAC clones BAC40E8, BAC31A8 and BAC64A5. **FIG. 37** is based on a phylogenetic analysis of the inferred amino acids of cloned proteorhodopsin genes. Evolutionary distances calculated from 220 positions were used to infer the tree topology by the neighbor joining method using the PaupSearch program of the Wisconsin Package version 10.0 (Genetics Computer Group (GCG), Madison Wisconsin). Other methods could also be used. The variants of the MB library share at least 97% identity over 248 amino acids, as shown in **FIG. 38**, and 93% identity at the DNA level. All the PCR amplified proteorhodopsin genes from Antarctic marine bacterioplankton (e.g. the PAL clones) were different from those of Monterey Bay (e.g. the MB clones) sharing 78% identity over 248 amino acids with the Monterey clade. The changes in amino acid sequences were not restricted to the hydrophilic loops, but spread over the entire protein including changes near the retinal binding domain **3830** as shown in **FIG. 38**, which are predicted retinal-binding residues. **FIG. 38** shows an example of a multiple alignment of proteorhodopsin amino acid sequences that were obtained from different clones **3820**. The secondary structure is derived from hydropathy plots (boxes **3810** shows trans-membrane helices).

Light-driven energy generator

FIG. 39 provides a light-driven energy generator **3900** that utilizes proteorhodopsin, as obtained from naturally occurring bacteria as described above, to convert light-energy into biochemical energy. Light-driven energy generator **3900** takes advantage of the

functional properties of the proteorhodopsin protein once expressed in, for instance, *E. coli* and other bacteria. These properties include the ability of proteorhodopsin 3906 to integrate within the cell membrane 3904 of, for instance, *E. coli* making an integrated proteorhodopsin protein 3908 (also called an integrated cell membrane protein). These properties also include the ability of proteorhodopsin 3906 to bind retinal 3910, making a light absorbing pigment 3912. The source of retinal 3910 is not limited to chromophore retinal but could also include chemical derivatives of retinal, such as 3-methyl-5-(1-pyryl)-2E,4E-pentadienal, 3,7-dimethyl-9-(1-pyryl)-2E,4E,6E,8E-nonatetraenal, all-trans-9-(4-azido-2,3,5,6-tetrafluorophenyl)-3,7-dimethyl-2,4,6,8,-nonatetraenal and 2,3-dehydro-4-oxoretinal. Illuminating light absorbing pigment 3912 with a light source 3914 results in a chemiosmotic gradient or proton pump in which light energy 3916 is converted into biochemical energy 3918. The chemiosmotic gradient involves pumping of protons from the inside to the outside of cell membrane 3904. When the protons return to the inside of cell membrane 3904 it produces biochemical energy 3918 via a proton translocating ATP-ase. Finally, the biochemical energy 3918 is harnessed by a mediator 3920 to produce energy 3922 for a particular process. For example, since proteorhodopsin functions as a light driven proton pump, it generates energy in the form of a proton motive force across the host cell membrane upon illumination. This light-driven proton motive force can be converted to many other forms of energy, one example above being the regeneration of adenosine triphosphate (ATP), via a proton-translocating ATPase. This coupling of the proton motive force generated by proteorhodopsin, for use by proton-translocating ATPases to synthesize ATP, could be accomplished both in living cells, as well as in artificially constructed membrane systems such as liposomes. Proteorhodopsin-based systems can convert light energy to a wide variety of useful mechanical, chemical, and electrical energy forms, for many industrial and technological applications. These

include, but are not limited to, use in targeted drug delivery, uses as primary or secondary energy generators for biocatalytic reactors, fuel cells and nano-machines (including molecular motors), as well as uses in molecular switching or data storage devices.

Applications that can potentially benefit from proteorhodopsin-light driven energy generation are, for instance, bio-electronics applications that are aimed to interface, integrate, or substitute the silicon based microelectronics systems as well as molecular devices. Other applications that can potentially benefit from proteorhodopsin-light driven energy generation are, for instance, in bio-materials, wherein proteorhodopsin is integrated as a bio-material in, for instance, optical films for light mediated computer memory applications, optical information storage and pattern recognition.

Alternatively, proteorhodopsin is useful for a process to enhance yield or increase the potential of recombinant protein production or converting the light induced membrane potential into cellular signals, including modulation of gene expression. The biochemical energy derived from functional proteorhodopsin exposed to light could be harnessed to support a variety of cellular processes. For instance, the energy derived from light-mediated proton pumping could be used to enhance the production of secondary metabolites, or recombinant proteins in host cells, such as *E. coli*. Often, production of specific compounds in the biotechnology industry is limited, since their optimal expression or production occurs in the late stationary phase of growth, when energy reserves of the host cells are low. Retinal-bound proteorhodopsin expressed in such cells would provide an ample source of biochemical energy, by simple illumination. Proteorhodopsin-mediated light driven proton production could enhance any variety of biosynthetic or physiological processes which require energy.

The biochemical energy derived from proteorhodopsin light driven proton pumping could also be converted to other generally useful energy forms, for example electricity. Microbial fuel cells currently use carbon-based compounds, such as glucose, as the primary energy source. Via specific mediators of reduction potential (e.g. electrons), these microbial fuel cells convert cellular biochemical energy to electrical potential. Unlike carbon-based microbial fuel cells, proteorhodopsin uses light as the energy source, that can then be converted into a chemiosmotic potential, and finally into cellular biochemical energy by membrane-bound proton ATP-ases. Therefore, the use of proteorhodopsin could be employed to derive energy from light as the primary or supplementary energy source, that could then be converted into electrical potential (analogous microbial fuel cells that derive their energy from glucose).

In addition to energy generation in vivo in living cells, membranes containing proteorhodopsin could be used to enhance or enable other specific processes in vitro. Polymers produced from proteorhodopsin-containing membranes may have specific properties that could be used similarly to those containing bacteriorhodopsin. One example includes the use of these light sensitive molecules for optical computing applications.

As shown in FIG. 39, the kinetics of proteorhodopsin as it is utilized in 3900 is influenced by various factors such as the type of light source 3914 and the manipulation of light source 3914 in terms of frequency and/or wavelength at which the light 3916 is delivered. Light source 3914 could be any type of light source that delivers light energy 3916 that would be absorbed by light absorbing pigment 3918.

For example, the light source 3914 could be tuned to optimally excite rhodopsin variances with an absorbance maximum of 490 nm or alternatively those rhodopsins with an absorbance maximum of 520 nm. Manipulation of the light source 3914 or the light 3916 being emitted by the light source 3914, for example, involves changing the frequency of fast-light pulses or the delivery of light 3916 as individual pulses, a train of pulses, or a continuous source of light. Manipulation also involves changing the wavelength of the delivery of light 3916 at different wavelengths. In addition, as is clear for one skilled in the art, changing the frequency and/or amount of retinal that will bind within integrated cell membrane protein 3908 also varies the function of proteorhodopsin. Finally, as was mentioned in the previous section, genetic variants of proteorhodopsin result in variants of the proteorhodopsin proteins that changes the kinetics of 3600 due to a difference in absorption of light at different wavelengths. The functional expression of such variation in these proteorhodopsin proteins adds another source of variation to the kinetics of proteorhodopsin as it is utilized in 3900.

As shown in FIG. 39, the light-driven energy generator includes a host 3902. In the present invention, as a preferred embodiment, host 3902 is a cell membrane preparation of *E. coli*. However, the present invention is not limited to the use of *E. coli* and, alternatively, other bacteria or eukaryotes could be used to provide host 3902 as an intact cell (in vivo) and/or as a cell membrane preparation (in vitro). For example, but not limited to, bacteria and yeast with developed genetic systems such as *Bacillus* spp. Species, *Saccharomyces* spp., *Streptomyces* spp. or *Pichia* spp. could be used as host for the expression of proteorhodopsin. In addition, in case a cell membrane preparation (in vitro) is used, host 3902 becomes equivalent to cell membrane 3904.

The light-driven energy generator 3900, as shown in FIG. 39, further includes proteorhodopsin 3906. Proteorhodopsin is presented in the form of the earlier presented expression vector containing a proteorhodopsin gene or one of its variants. Once proteorhodopsin 3906 has been put into host 3902, the proteorhodopsin expression vector expresses the proteorhodopsin protein in host 3902. An integral cell membrane protein 3908 is created in which the proteorhodopsin protein inserts into and folds properly within the cell membrane 3904. This is accomplished in the *E. coli* host by virtue of the native signal sequence found in the 5' end of the proteorhodopsin gene. It could also be accomplished by replacement of native sequence with another host-specific signal sequence in non-*E. coli* host systems.

As shown in FIG. 39, once retinal 3910 is added to cell membrane 3904, retinal 3910 binds within integrated cell membrane protein 3908 and forms a light absorbing pigment 3912. The particular example of FIG. 40 shows an integrated proteorhodopsin protein 3908 bound to retinal 3910 in *E. coli*. Chemical derivatives of retinal (as discussed above) could also be used as a substitute chromophore to generate functional proteorhodopsin. For the particular example of FIG. 40, the proteorhodopsin protein was cloned with its native signal sequence and included an addition of the V5 epitope, and a polyhistidine tail in the C-terminus. The proteorhodopsin protein was expressed in host 3902, i.e. *E. coli* outer-membrane protease-deficient strain UT5600, and induced with 0.2 % arabinose for 3 hours. Cell membranes 3904 were prepared and resuspended in 50 mM Tris-Cl (pH 8.0) and 5 mM MgCl₂. FIG. 40 shows a proteorhodopsin-expressing *E. coli* cell suspension. After 3 hours of induction in the presence of 10 µM all-trans retinal, cells expressing the protein acquire a reddish pigmentation as indicated by 4010 and the + (plus) symbol. FIG. 40 also shows that a cell suspension using the same PCR primers

(Sequence ID No:2 and 3) but now in opposite orientation as a negative control, did not acquire a reddish pigmentation as indicated by 4020 and the – (minus) symbol.

FIG. 41 shows an exemplary embodiment of the absorption spectra of light absorbing pigment 3912 upon illumination with light source 3914 as is shown in **FIG. 39**. As mentioned above, the light absorbing pigment is a retinal-reconstituted proteorhodopsin in *E. coli*. **FIG. 41** shows absorption spectra of light absorbing pigment 3912 as well as a negative control. After retinal 3910 addition to integrated proteorhodopsin protein 3908, light absorbing pigment 3912 was made. The retinal 3910 addition was done at selected time points, i.e. 10, 20, 30 and 40 min, and shows a progression from low to high absorption values indicated by respectively 4110, 4120, 4130 and 4140 upon illumination with light source 3914. **FIG. 41** also shows the absorption spectra of retinal 3910 addition at these similar time points but now to a negative control of retinal 3910 containing a proteorhodopsin 3906 that was created using the same PCR primers in opposite orientation. 4150, 4160, 4170 and 4180 indicate the four absorption spectra for the negative control. An absorption peak at 520 nm was observed after 10 minutes (4110) of incubation as illustrated in **FIG. 41**. On further addition of retinal, the peak at 520 nm increased, and had a ~100 nm half bandwidth. The 520 nm absorption peak was generated only in membranes containing proteorhodopsin 3906, and only in the presence of retinal 3910. The red shifted λ_{max} of retinal ($\lambda_{\text{max}} = 370$ nm in the free state) is indicative of a protonated Schiff base covalent linkage of retinal to proteorhodopsin.

FIG. 42 shows an exemplary embodiment of the light mediated proton pump of the light-driven energy generator 3900 indicating the conversion of light energy 3916 as shown in **FIG. 39**. The proton pump action is illustrated by measuring pH changes in

the medium surrounding the host 3902, which in this particular example involves a cell suspension of *E. coli*, illuminated by light source 3914. The beginning and cessation of illumination (with yellow light >485 nm delivered by 3916) is indicated 4110 ("ON") and 4120 ("OFF") respectively. The cells were suspended in 10 mM NaCl, 10 mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ and 100 μM CaCl_2 . Net outward transport of protons was observed solely in proteorhodopsin-containing *E. coli* cells, and only in the presence of retinal 3910 and light 3916 and is indicated by 4210 in FIG. 42. Light-induced acidification of the medium was completely abolished by the presence of 10 μM of the protonophore CCCP.

FIG. 43 is an exemplary embodiment showing that illumination by light source 3914 generates an electrical potential at the membrane 3904 in proteorhodopsin-containing right-side-out membrane vesicles, in the presence of retinal 3910, reaching -90 mV after 2 minutes from light 3916 onset. Transport of $[^3\text{H}]\text{TPP}^+$ in *E. coli* right-side-out vesicles containing expressed proteorhodopsin, reconstituted with (4310 and 4320) or without (4330 and 4340) 10 μM retinal 3910 in the presence of light (4310 and 4330) delivered by the light source 3914 or in the dark (4320 and 4340). FIG. 43 shows that proteorhodopsin, in its form of 3912 as a light absorbing pigment, pumps protons from the inside to the outside of cell membrane in a physiologically relevant range. The ability of proteorhodopsin to generate a physiologically significant membrane potential, even when heterologously expressed in nonnative membranes, is consistent with the proton pumping function for proteorhodopsin in the native gamma proteobacteria from which it is derived.

FIG. 44 is an exemplary embodiment showing that proteorhodopsin can have a fast photocycle and can therefore be characterized as a fast and therefore efficient

transporter of protons. For the particular example of FIG. 44, light absorbing pigment 3912 is induced by laser pulses delivered by light source 3914. Laser pulse-induced absorption changes are shown by 3912 in host 3902, which in this case are suspensions of *E. coli* membranes containing proteorhodopsin. A 532-nm pulse (6 ns duration, 40 mJ) was delivered at time 0 and absorption changes were monitored at various wavelengths in the visible range in a lab-constructed pulse photolysis system. 64 transients were collected for each wavelength. 4410 indicates transients at 3 wavelengths exhibiting maximal amplitudes. 4420 indicates absorption difference absorption spectra calculated from amplitudes at 0.5 ms (indicated by 4430) and between 0.5 ms and 5.0 ms (indicated by 4440). In 4410, transient depletion occurred near the absorption maximum of pigment 3912 (500-nm trace indicated by 4450), and transient absorption increase was detected at 400 nm (indicated by 4460) and 590 nm (indicated by 4470), indicating a functional photocyclic reaction pathway. In 4420, the absorption difference spectrum shows that within 0.5 ms an intermediate with maximal absorption near 400 nm is produced (indicated by 4430), typical of unprotonated Schiff base forms (M intermediates) of retinylidene pigments. The 5-ms minus 0.5-ms difference spectrum 4440 shows that following M decay an intermediate species red-shifted from the unphotolyzed 520-nm state appears. The decay of proteorhodopsin final intermediate is the rate limiting step in the photocycle and is fit well by a single exponential process of 15 ms, with an upward baseline shift of 13% of the initial amplitude.

As mentioned above, a proteorhodopsin gene or protein variant can be selected to determine an absorption spectra of the light absorbing pigment to change the kinetics of the light energy generator 3900, for instance to meet a design/functional criteria of an application wherein proteorhodopsin is utilized. FIG. 45 shows an exemplary

embodiment of different absorption spectra of retinal-reconstituted proteorhodopsins in *E. coli* as a function of wavelength 4510. As shown in FIG. 45, the absorbance 4520 is different and depends on the clone from which the proteorhodopsin was amplified. In this particular example, 5 μ m all-*trans* retinal was added to the membranes suspensions in a 100 mM phosphate buffer, with a pH 7.0, and absorption spectra were recorded. The four spectra 4530, 4540, 4550, and 4560 are respectively for the proteorhodopsin genes retrieved from clones HOT75m4, PALE6, HOT0m1, and BAC31A8 at 1 hour after retinal addition. The proteorhodopsin gene retrieved from clone HOT75m4 4530 and PALE6 4540 produced a blue (490 nm) absorption maximum. The proteorhodopsin gene retrieved from clone HOT0m1 4550 and BAC31A8 4560 produced a green (527 nm) absorption maximum. In general, a range of wavelengths could be obtained that is not limited to the range shown in the example of FIG. 45.

It will be clear to one skilled in the art that the above embodiment may be altered in many ways without departing from the scope of the invention, such as for instance by mutagenesis to change the genetic sequence of proteorhodopsin and thereby changing the kinetics of the proteorhodopsin protein once it is expressed. Accordingly, the following claims and their legal equivalents should determine the scope of the invention.

DEPOSITS

Depository address: 10801 University Boulevard, Manassas, VA 20110, USA.

The *Escherichia coli* containing cloned DNA BAC 31A8 having assigned ATCC number PTA-3083, the *Escherichia coli* containing cloned DNA BAC 40E8 having assigned ATCC number PTA-3082, the *Escherichia coli* containing cloned DNA BAC 41B4 having assigned ATCC number PTA-3080, and the *Escherichia coli* containing cloned DNA BAC 64A5 having assigned ATCC number PTA-3081, all having been deposited on February 21, 2001 with the ATCC Patent Depository.

The *Escherichia coli* containing a plasmid PAL E6 having assigned ATCC number PTA-3250, the *Escherichia coli* containing a plasmid HOT 0m1 having assigned ATCC number PTA-3251, the *Escherichia coli* containing a plasmid HOT 75m4 having assigned ATCC number PTA-3252, and the *Escherichia coli* containing cloned DNA BAC64A5 having assigned ATCC number PTA 3082, all having been deposited on March 30, 2001 with the ATCC Patent Depository.

LIGHT-DRIVEN ENERGY GENERATION USING
PROTEORHODOPSIN

LIST OF SEQUENCES THAT ARE LISTED IN THE
INCORPORATED SEQUENCE LISTING

- Sequence ID No:1** bacterial artificial chromosome (BAC) clone 31A8 (EBAC31A8).
- Sequence ID No:2** nucleotide sequence of proteorhodopsin-specific polymerase chain reaction (PCR) primer 1.
- Sequence ID No:3** nucleotide sequence of proteorhodopsin-specific polymerase chain reaction (PCR) primer 2.
- Sequence ID No:4** nucleotide sequence of the proteorhodopsin gene amplified from clone EBAC31A8 (Sequence ID No. 1) using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:5** deduced amino acid sequences of the proteorhodopsin gene amplified from clone EBAC31A8 (Sequence ID NO:4).
- Sequence ID No:6** native proteorhodopsin nucleotide sequence from clone EBAC31A8 (Sequence ID No:1).

- Sequence ID No:7** deduced amino acid sequences of the native proteorhodopsin nucleotide sequence from clone EBAC31A8 (Sequence ID No:6).
- Sequence ID No:8** nucleotide sequence of the proteorhodopsin gene amplified from clone EBAC40E8 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:9** deduced amino acid sequences of the proteorhodopsin gene amplified from clone EBAC40E8 (Sequence ID NO:8).
- Sequence ID No:10** nucleotide sequence of the proteorhodopsin gene amplified from clone EBAC41B4 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:11** deduced amino acid sequences of the proteorhodopsin gene amplified from clone EBAC41B4 (Sequence ID NO:10).
- Sequence ID No:12** nucleotide sequence of the proteorhodopsin gene amplified from clone EBAC64A5 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:13** deduced amino acid sequences of the proteorhodopsin gene amplified from clone EBAC64A5 (Sequence ID NO:12).

- Sequence ID No:14** nucleotide sequence of the proteorhodopsin gene amplified from clone HOT0m1 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:15** deduced amino acid sequences of the proteorhodopsin gene amplified from clone HOT0m1 (Sequence ID NO:14).
- Sequence ID No:16** nucleotide sequence of the proteorhodopsin gene amplified from clone HOT75m1 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:17** deduced amino acid sequences of the proteorhodopsin gene amplified from clone HOT75m1 (Sequence ID NO:16).
- Sequence ID No:18** nucleotide sequence of the proteorhodopsin gene amplified from clone HOT75m3 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:19** deduced amino acid sequences of the proteorhodopsin gene amplified from clone HOT75m3 (Sequence ID NO:18).
- Sequence ID No:20** nucleotide sequence of the proteorhodopsin gene amplified from clone HOT75m4 using PCR primers according to Sequence ID No:2 and Sequence No:3.

- Sequence ID No:21** deduced amino acid sequences of the proteorhodopsin gene amplified from clone HOT75m4 (Sequence ID NO:20).
- Sequence ID No:22** nucleotide sequence of the proteorhodopsin gene amplified from clone HOT75m8 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:23** deduced amino acid sequences of the proteorhodopsin gene amplified from clone HOT75m8 (Sequence ID NO:22).
- Sequence ID No:24** nucleotide sequence of the proteorhodopsin gene amplified from clone MB0m1 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:25** deduced amino acid sequences of the proteorhodopsin gene amplified from clone MB0m1 (Sequence ID NO:24).
- Sequence ID No:26** nucleotide sequence of the proteorhodopsin gene amplified from clone MB0m2 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:27** deduced amino acid sequences of the proteorhodopsin gene amplified from clone MB0m2 (Sequence ID NO:26).

- Sequence ID No:28** nucleotide sequence of the proteorhodopsin gene amplified from clone MB20m2 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:29** deduced amino acid sequences of the proteorhodopsin gene amplified from clone MB20m2 (Sequence ID NO:28).
- Sequence ID No:30** nucleotide sequence of the proteorhodopsin gene amplified from clone MB20m5 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:31** deduced amino acid sequences of the proteorhodopsin gene amplified from clone MB20m5 (Sequence ID NO:30).
- Sequence ID No:32** nucleotide sequence of the proteorhodopsin gene amplified from clone MB20m12 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:33** deduced amino acid sequences of the proteorhodopsin gene amplified from clone MB20m12 (Sequence ID NO:32).
- Sequence ID No:34** nucleotide sequence of the proteorhodopsin gene amplified from clone MB40m1 using PCR primers according to Sequence ID No:2 and Sequence No:3.

- Sequence ID No:35** deduced amino acid sequences of the proteorhodopsin gene amplified from clone MB40m1 (Sequence ID NO:34).
- Sequence ID No:36** nucleotide sequence of the proteorhodopsin gene amplified from clone MB40m5 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:37** deduced amino acid sequences of the proteorhodopsin gene amplified from clone MB40m5 (Sequence ID NO:36).
- Sequence ID No:38** nucleotide sequence of the proteorhodopsin gene amplified from clone MB40m12 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:39** deduced amino acid sequences of the proteorhodopsin gene amplified from clone MB40m12 (Sequence ID NO:38).
- Sequence ID No:40** nucleotide sequence of the proteorhodopsin gene amplified from clone MB100m5 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:41** deduced amino acid sequences of the proteorhodopsin gene amplified from clone MB100m5 (Sequence ID NO:40).

- Sequence ID No:42** nucleotide sequence of the proteorhodopsin gene amplified from clone MB100m7 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:43** deduced amino acid sequences of the proteorhodopsin gene amplified from clone MB100m7 (Sequence ID NO:42).
- Sequence ID No:44** nucleotide sequence of the proteorhodopsin gene amplified from clone MB100m9 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:45** deduced amino acid sequences of the proteorhodopsin gene amplified from clone MB100m9 (Sequence ID NO:44).
- Sequence ID No:46** nucleotide sequence of the proteorhodopsin gene amplified from clone MB100m10 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:47** deduced amino acid sequences of the proteorhodopsin gene amplified from clone MB100m10 (Sequence ID NO:46).
- Sequence ID No:48** nucleotide sequence of the proteorhodopsin gene amplified from clone PALB1 using PCR primers according to Sequence ID No:2 and Sequence No:3.

- Sequence ID No:49** deduced amino acid sequences of the proteorhodopsin gene amplified from clone PALB1 (Sequence ID NO:48).
- Sequence ID No:50** nucleotide sequence of the proteorhodopsin gene amplified from clone PALB2 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:51** deduced amino acid sequences of the proteorhodopsin gene amplified from clone PALB2 (Sequence ID NO:50).
- Sequence ID No:52** nucleotide sequence of the proteorhodopsin gene amplified from clone PALB5 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:53** deduced amino acid sequences of the proteorhodopsin gene amplified from clone PALB5 (Sequence ID NO:52).
- Sequence ID No:54** nucleotide sequence of the proteorhodopsin gene amplified from clone PALB7 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:55** deduced amino acid sequences of the proteorhodopsin gene amplified from clone PALB7 (Sequence ID NO:54).

- Sequence ID No:56** nucleotide sequence of the proteorhodopsin gene amplified from clone PALB6 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:57** deduced amino acid sequences of the proteorhodopsin gene amplified from clone PALB6 (Sequence ID NO:56).
- Sequence ID No:58** nucleotide sequence of the proteorhodopsin gene amplified from clone PALB8 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:59** deduced amino acid sequences of the proteorhodopsin gene amplified from clone PALB8 (Sequence ID NO:58).
- Sequence ID No:60** nucleotide sequence of the proteorhodopsin gene amplified from clone PALE1 using PCR primers according to Sequence ID No:2 and Sequence No:3.
- Sequence ID No:61** deduced amino acid sequences of the proteorhodopsin gene amplified from clone PALE1 (Sequence ID NO:60).
- Sequence ID No:62** nucleotide sequence of the proteorhodopsin gene amplified from clone PALE6 using PCR primers according to Sequence ID No:2 and Sequence No:3.

Sequence ID No:63 deduced amino acid sequences of the proteorhodopsin gene amplified from clone PALE6 (Sequence ID NO:62).

Sequence ID No:64 nucleotide sequence of the proteorhodopsin gene amplified from clone PALE7 using PCR primers according to Sequence ID No:2 and Sequence No:3.

Sequence ID No:65 deduced amino acid sequences of the proteorhodopsin gene amplified from clone PALE7 (Sequence ID NO:64).

ttggttatatc agtaaatgggt attggtccaa taacttaata ctaatatata attagtttat 60
gaataaaattt tatatatattg ggttattgtt ttttacta aatgcatttt cttgctcaga 120
tcttctagat acagacatga gagtcttga tccgctgag tcaagaaacc tttgcgagtt 180
tgaaggaaaa gctttactag ttgtgaatgt tgcaagtaga tgtggttaca cttatcaata 240
tgctggcctt caaaagtatt atgaaagtta taaagatgaa gattttcttag taattgggat 300
cccatctaga gatttttcttc aagaatactc tgatgaaagc gatgttgcag aattttgttc 360
tacagaatac ggtgtttgaat ttcctatgtt ctcaactgct aaagtcaaag gaaaaaaagc 420
acaccattt tataaaaaac ttattgcaga atcagggttt actccctcat ggaaactttaa 480
taaatactta atctcaaaag agggcaaggt tgtatccaca tatggatcaa aggtaaaagc 540
tgattcaaaa gagcttatat cagctataga aggcttgctg taaaattatt acttagaaac 600
taatacagtt ttaggcttgt ttgctgcaaa tattccatta tctacaactc caggaatat 660
attaatcaaa gcttccattt cagtggggtt tgaatatatcc atattagaga tatctaaaat 720
gtgattacct tgggtctgtta taaatccagt tctatatgtg ggtattccac cgatcgagat 780
tatttttctt gcaacaaggc tcctacttcc aggtatcacc tctataggca gtggaaaagc 840

tcccaaaaga ttaaccatct ttgactgac aactatacat ataaactcgt tagaggcaga 900
agcaactatc ttttctctag tatgtgcgcc accaccacct ttaataagac aattttcagg 960
agacacctca tctgcaccat ctatgtaata agctatatca actacatcat taaggctaaa 1020
gacctctatc ccattttcat ttaataattt tgatgaagca tctgaactag aaacagctcc 1080
agcaaatgtg tgcctatgct ccttagttc ttctataaaa aaattaactg ttgagccggt 1140
tccaatacct aaaatcatct caggatgaag attatttttg atataattcta tagcttggtt 1200
agcaacattt atctttgagc cactcataga gttataatac aagaaaaatat aggtagttaa 1260
ttattttgag actaaaaaatt aaaaaaacag gttcttttaa gaattcccag aagtacctaa 1320
agcttattcc taatgcatct gtttatgacg tagcattaaa atcacctata acatttgctc 1380
taaatatttc ttcaaagctg gggaataaag ttttcctaaa aagagaggat ctgcaacctc 1440
tattttcttt taaaaacaga ggagcgtata acaagattgt aaatttatcc gatgccgaaa 1500
agaagagggg ggttattgct gcatacagcag gaaatcatgc tcaaggggta gccagtgc 1560
gtaagaaaatt aaaaattaat tgcttgatag ttatgccaat acaactcca gaaataaaaa 1620
taaaagatgt aaaaagattt ggagccaaaa tactccaaca tggggacaac gtagatgcag 1680

cattaaaaga ggcactgttt attgcaaaga aaaaaaatt gtcttttgtt catccttttg 1740
acgaccctct aacaattgct ggccaaggga ctataggaca agaaattctt gaagataaaa 1800
ataaatttga tgttgtcttt gtccggttg gaggaggagg tattctagct ggtgtatctg 1860
cctggatagc acagaataat aagaaaataa aaattgttgg tgttgaggtt gaggattccg 1920
cttgtcttgc tgaggccgta aaagctaata aaagagttat tttaaaagaa gtgggcctct 1980
ttgctgatgg ggtggcagta tcaagggttg gaaaaataa ttttgatgtt attaaagagt 2040
gcgtagatga agtcattaca gttagcgttg atgaggcttg caccgctgta aaagatatct 2100
ttgaagatac aagggttcta tcagaacctg ctggggcatt agcacttgca gggttaaaag 2160
cctacgcaag gaaagttaaa aataaaaaac ttattgctat aagtcttggc gctaatgtaa 2220
atttccaaag acttaattt attgttgagc gatcagagat tggtgaaaaat agagaaaaaa 2280
tattaagtat caaaatccca gagataacctg gaagttttct taagctttca aggatgtttg 2340
gcagctctca agttacagag tttaactaca ggaaatctag cttaaagcgt gcatatgttt 2400
tagttgggtgt tagaactaaa actgaaaaat catttgaaat cttaaagtc aaattaaaaa 2460
aagcaggctt caccttttagc gacttttact gaaatgaaat atccaatgat catctgaggc 2520

atatggttgg tggcagaaat agtgactcag gctctcataa caatgaaaga atatttaggg 2580
gagagtttcc tgagaagccg ggcgcgctgt taaattttct agagaaattt ggaaataaat 2640
ggmatatttc cttatttcac tacaggaacc taggttcagc ttttggaag atattaattg 2700
gcatcgagag taaggataaa gacaagctaa taaatcattt aaataagtca ggnactattt 2760
ttacagaaga aacctctaac aaggcataca aagattttttt aaaatgaaag gttataactt 2820
taatctaaat ttaattgaaa aaagctcacc gctagggtttt tcccacggct ctttgaacaa 2880
ctcggattga gatctatcat cctcctcgtc gtaaattctc ccacctttag aatagaccaa 2940
aaatagatat gacaaaggag cgagctcata tttatatcta atttggaacg aagctacgcc 3000
agtattaaat tcattaacta tattatttcc tttataaagg tatccattag catctgaaaa 3060
aatactaata ggatttttttg ctttcaaagc aacaaaattga ctcttaagtc tgatctcatg 3120
tttattattt ttaaaccaat ttagatcaaa agataaggta tcttggtcttg aatcatatga 3180
ggcaagatta ttattatcct gccatatcag ccattcattt tcttttctta ttctatatg 3240
cgcattgatt cttaagttat catttggaag tattgaacct gctatcttgt aaaattttct 3300
accataccca ttcgaatccc accgattatc tttctctccc ttaaagaagc taactctcca 3360

gtcatatgtc cagaatgaat agttctttgc ctcaaagtct gctgtaatac ctattcttct 3420
ttttgatttg ataaaggggt aggccttcatt ttttcttgtg atagttgtat ttttcccaga 3480
agatctaaag ttaaaatcta attgaaattt agagttgtcc ttaaaaactaa aagaattttt 3540
ttgatcgatg cctattggat ttgaattacc agctgtgtca gcatcataat ttagatcaat 3600
tccataatct atttgtttta atatcgagct attatcaaat tcatttattt ttcgattttcc 3660
tccaatacca gcatgaatcc agtctcttct ttgcagataa ccaaagtcac ttaactcgaa 3720
gtcgtcttca aaataaaagaa ggcttccact tatgttagat agtttatattg gaagatatgt 3780
aaactgagtc ctatacccaa gccattttt accatctttt tctgaggcta acagatctga 3840
atatgtaatt aatttttttg aacgaatatt gatgtaatca ataacattga ctgttgatga 3900
ttcgcctgtc atttcattct caacattcgt taccatgaag ccaagcgtt tatttccaag 3960
ctttgttcga gatcgaagag cataatagtc tcttccaact gaaaaggctt catcagcctc 4020
acttgctaca aatactccaa attcattatt attacttttt tgagtaagtc ttaatgcaaa 4080
atcaatatca gaatagtttt ttttagctgc ctgcgaacct tcttcattac tttcttctga 4140
gcaattatag ctggggggcag ctccaatcct ccgcgtattt ataaccgagt atctatcata 4200

attactaata tcaaatagtg attgggttttc attgaaaaat gctctttttt ctgagtaaaa 4260
agtttcttga gcagaaaaagt taataaccac atcatcactc tcagcttgtc cgaaatctgg 4320
attaatagct aaatttattt gacgacctt tccagtgtct taaaagattt cagccccaat 4380
atctgaccct tcttggtttg taactgaatt tttatttgaa gatatatatg gaaaaaaagt 4440
aagctttgat tttgtatagt tttgtatttc taagctatct aactcttgaa agtagtcatt 4500
tctactagcc attgttccgg cactgctaac ccatgactca ttttcggcac tataacgtaa 4560
tgcggtgtaa ttaatttttc ttatatcacc atcaggctgt ttcatthaacg ttacatccca 4620
aggaataaaa aactcagaga cccaataccc atcaaatttt tgtgtttttg caatccaatc 4680
tccatcccag tctgtcttaa agtctcctgc ttgcgttttt atggcatcga aaagcgagtt 4740
cccaagattt atagcaagaa tgaaaagctt gtaccatca ccatcaaat ctatatatat 4800
agagttttta tcgctaagtg aatttatttg atctctaagc gtcttcctgg agaacaataga 4860
atcattactt tgaaaattct taaatccaac atatatacca tccttatttg agaaaaattaa 4920
agccgtttga agaagttcat ttttcttaag agtaaaaagga gatgtttcat aaaaatctgt 4980
aatttcaaat gcattattcc actcaggctc atcaagagag ccatcaataa caattgagtt 5040

tgaccaaatc agcacagagg taagtaaaag tgataatgag gctaaagtta ttttcataga 5100
tagatttaat tcagagtatt ttaatctatg aagagtagga aatcatccga tcattctcag 5160
aaaaaacata atcatattta aggtcatggt tttctccaaa atgcataatta caaatctggt 5220
attgataaca cagtccaacc aataaaggtc ttgatacctc ctcgtttata gagccaataa 5280
ttctatcgaa atatccagag ccatagccaa gcctgtatcc attttaatca actcctgtca 5340
taggaataaa cattaaatca atttcattta tgttgacata atcttcactt ttaacctctt 5400
tgatcccaaa ttgatttata aagaagttag gctgctcadc cagcaaatca aaagccatca 5460
tttcgtcadc aattactttt ggaatataaa tattcttttt aagttagta aaagcttgaa 5520
ttaacaaatt tgtattgact tcatttcgaa aaggaaaata taaggcaatg ttttgcat 5580
catgagtatt gattttttct aatacat 5640
ataaatcaga aatagactgc ccctgctcaa agagtgattt tcgtatttta tttttcacca 5700
taaatgtgct cgaggctaac aaaccaaaac accgccatgt gctataacctg aaccgatggg 5760
tcagggtgtg aatttcataa caaatcaggc ttccctgttaa aggtactgag caacaatgct 5820
agaggaatat cacctattta attgtatcgg ttcaaaattta ttaacacatt agcgtatgaa 5880

ccagaatgca gttaattata aaatatatag gtattaagta aaagttaatt tttagagagc 5940
agactctatt ttttgtatta gcttttcgat atctttatta tttaaagcag agtcagtatt 6000
ctcggttggc ttggataaaa gcttactagc catagttaga ccagctatca ccaaggcatt 6060
attcttatca ctgatgccat caagctcatc atttaataat ttgcagctc taatgagttg 6120
gtctttctct tctggcggac aggctaactg cagatctcgg ccaaaaattc ttaaagatag 6180
cgtttccatt tctgacatta ttttttgagt ttagttattt ctttatttag tatgagctta 6240
tcttttttcc agtctctttc cctttgcaca taagaatcta ttattccttt ttgctcatca 6300
aatcttttta tcaaaagatc taccttatcc tctagttcaa ataatgatga ttttttttct 6360
tctgacattg tataagttta attgtactag aatgaattgg aaagtttctt taataaaaga 6420
gtaaaataag gcatggaaaa aataattttt aaaaacagaa gagattcact tataaaacac 6480
ctacctaaag attcagcctt aatgtacct ggtgcagatt tgcaatatag aaatgctgac 6540
tcattcttata atttaagaca ggaaagtagc ttctactatc tgtctggcctt ttgtgagccc 6600
tcttctctaa tggtttttagt taataatgga aaagcattg attcaataat ttttgttcct 6660
gaaaaagata aacttaaaaga aatctgggat ggttatcgag ctggccctga gggcgcaata 6720

aatgattttc tttttgatca agctttttgaa aataataaat cagatgcttt aatgcctgaa 6780
atccttcaag ggctagaaaa agttttttat tcaatagga agaaaaatgg ctttgatcag 6840
aaagtaattg actggacatg cgcagcaaat tctaaagata ggcacagcaa atcaattgat 6900
attattgatg gctcttcgat ggtaggaat ttaaggctta tcaaagataa gcatgaaatt 6960
gatattatga agagagcttg tgaaaatttca gctgaatcat atattgaggt catgaaatct 7020
ataaagcctg gggacaatga gcaggaaata gaggcgctat ttttatatga attcgccaaa 7080
agggaggaa ggtttccagc ttatacacct atagtgtctg gaggtgaggg tgcttgtgta 7140
ttgcattata ttgaaaaatga taaagagtta gcttcatcag atttaatttt ggtagacgca 7200
ggatgtgaat acaaaaatgta tgcattctgat atcacaaaga ccttcccagt aagtggaaaa 7260
ttttcagatg aacagctaca aattataat attgtccaca aagccaatct tgctgcaatc 7320
gatgctgtaa aaactggtaa tagcataatg gagccccaaa tggtttcaga aaaagtaatt 7380
actgaaggtc ttgtagagtt gggatttcta tctggcgatg ttaatcagct tcataaaaaat 7440
ggtgcattca aggactttta tatgcataag gtgggacatt ggcttggact tgatgttcat 7500
gatgttggtg actacatgga gggagatgag tttatgaagt ttaagccagg gatgataacc 7560

acaatcgagc caggcatcta tatcagtagc gcaatggatg tagatgacaa atggaaaaggc 7620
atcggcataa gaatagagga cgacatccctt gtaacagatt caggcaatat taatctaaca 7680
gagaagggtgc catctaatacc tcaagaaaata gaatcattga tggcttagac tatggagggtt 7740
ccaattgtta tttctggcgg aggataata ggtaattaca tttctcttag gcttgaaaaa 7800
aataatatca aaaccgttat tgtcgaaaaa gctagtagtt tcaaagccct agataagggt 7860
ataagaacag tcaactctcaa tgagcattct atgcaaatgc taaaaaatat tggatatgtc 7920
ccatcaattg ctcaaatcaa cagcatcgac gtattagatg gtgagggtac aggcaaaatt 7980
caatttctag caaaggacgt aggcagcga aacctttcat atgtaacctt tttcaatgaa 8040
ttacaaaaac taatttctga tccatgtaaa gaaagaacct tatttaataa tgagattgat 8100
tcagttcaga atcttaatac agaattctgat ccagagatca tgcttaaaga tggcatgacc 8160
ataaaaacga atctaattgc tggatgtgat ggaagaaaatt caaatattgc aaaaattgct 8220
tcacttacaa gcagcttcga tgactactta caaacagctt taacttttgt cgttgatatt 8280
gataatgatt cacatggcaa agctcaccaa gttttttctg aaaaaggaat atttgcaatt 8340
atgccactcc cagaaggcaa gggtagatg aataaatgca cagtggttg gtcaataaaa 8400

aatcaagttt tgggagatga gcctgtatct gagtttgtaa aaaataacat ttctttttttt 8460
gaatcaaagc ttaatgttag tctcaggggtt aagtcagaaaa ttttaagttt taaattatcg 8520
aaccatcatt ttgaaaaacta tattagcgga cctattgttc ttcttgggtga tgctgctcac 8580
tcaattcacc ccttagcagg tcaaggtatt aatctaggat ttgcagatgc agatactttt 8640
tgtgaaagagg taattagtct ttataaaaaa gggattgcct ttaatgagaa atcagttttta 8700
aaaagatatg agattagaag aaaaagtatg aacttttttaa tgttgaagtc tatggactttt 8760
tttgtggatt tattttggttc agaaaaattta tatctaaggc tgatcagaaa tttaggtatt 8820
tcttcggttaa ataaatcaaa gtttgtttaa gcatttcttta taagacatgc ctctggaatg 8880
aataagtttt aaatttgtat taaacttttt gaccttttagc tctaagtctt ttaagaaactt 8940
cactaatgcc ttttttatca atgattctca taccttttgc agatacttta agatttacga 9000
acctgttctc agattcaacc caaaatttgt gtgtgtgaag attaggaaaa aactttcttt 9060
tagtcctatt tttagcgtga gaaacattgt ttcttgactg tggatatctta cctgttactt 9120
gacatatatt actcattgaa acgcgatttt atagaacact gaggaactta gcaatactat 9180
tgtgaaacaa atttatttat tacggcatgc acaatctgac tgggagagct ctaatcagaa 9240

agattttgat agaccattag caagaaaagg cattgaagaa gcaaatataaa tatcatgtta 9300
ctgcaaatct cattcaattt tagtagataa aatatcttgt agcactgcag aaagaactaa 9360
gcagactttt gatatatgca gtgatgggct taattatcca atagctgaag cagtctatac 9420
tgatgagctt tacttttctg gccctgggtga gatcgttaag cttatcccaa gtttaagtga 9480
attcatttcc tctgtttttaa taataggcca caatccatca atgcaaatgt acatagatgc 9540
tatttcagaa aatcctcata ttacgtattc aacatgcggg ctggcagaaa ttctcgttga 9600
aagttcatgg aaagacttat ctttaaaaaa atgtaagtta aaatctttta ttcaaccagg 9660
agagctttaa aaattgaaaa acataaaaaa taagatcatt aaccactaa tgggatccaa 9720
gataccctta cctcaatatg aaacaaaagg ctcggcagga ttggatttaa gggcatgcct 9780
agatagtaat ctcagccttc aagcaggaac atctcagttg atacctattg gtttttgcaat 9840
gtacttagaa gatccccggtc ttgcagcaat gggtatacct agatcagggt taggtttctaa 9900
gcatggaatc gttctttggta atctggttgg gttgattgat tcagactatc aaggagagct 9960
aatggttcct gcctgggaata gatcagatac agatttttgag attaatacctg gagacaggat 10020
tgcacaaatg attatagttc cagtgattca agcagatttt gaaattgtag acgagttcaa 10080

tgagactcag aggggagaaa agggtttttg aagttcaggt ataaattgat aaatttactt 10140
tttctttgcca aatcttttctt caaattttctg aactctttccg ccagtatcaa taatttttttg 10200
cttaccggta taaaaaggat gagaagcaga ggatatatca agaggtagt atgggtatgt 10260
ttttccatct tccattctt tcgtttgagt cgtatctaata gttgaacgaa tgagaaaagaa 10320
cttatcagca ctacgtcat ggaataaaac ttcacgggtat tcaggatgta tatctttttt 10380
cataataaaa tcaaaatttg gatgagaact ataacaaaaa acaacttctt ttcaatcaaa 10440
acatgaaaaat tttttactat gacatagctg tttcatgtgc tctaaggcaa tgctttactt 10500
atagtttctga gcttaaaatt acaaaaggaa cccgcgtgtc agttcctttt ggtaaaaagga 10560
agattgtagg agtggttata aaaaatatc aaaagccaga tttcctaaaa aaagcggag 10620
ctataaaaaa aattatcgct gtacttgatg aatatacctt gtctgacaag ccaatttttg 10680
attctatatt gtggtcttct gattactatc atcatcctat tggtagggtt ttaataacct 10740
ttataccccc cgaattaaga aaaattaata ataaaaaaat tgaagcttta agagaatttt 10800
ctgaatatc agtaaatgag gatgataaga aattcgattt aaccaaggat caagaaaaag 10860
cagtcaaggc gcttttctaaa tctaaaggat tttcaccac tttattatat ggagttacag 10920

ggtctgga aacagaagtt tacttaagag ttgccgaac ttttattaaa aataataagt 10980
cagtattagt tttagttcca gaaataaatt taacccccca attgctctcg cgatttgaga 11040
atagatttaa tgggtgaaatt ggcatatatc attctaagca aacagcagct aagagattaa 11100
agacttggt aaaagctaaa tttggttcta taaaaataat agtaggaact cgatcttcgg 11160
ctttagtgc tttagataac attggtttaa taattatcga tgaagagcat gaccaatcat 11220
ttaggcagtc agaagggttt aaattctctg ctagagactt aagtataaaa agggcacagc 11280
ttgcagatat tccaattatt ttgggatcag caacccttc gctgcaaaact ttaaaacttg 11340
taaaagaaaa taaatttata agagttgata ttctaatcg agttgatga aacaagcctc 11400
ctaaattaat agccttagat atcaataaca gccctttaat aggcggagtt gctaaagaga 11460
caattgaagc aatgcaatca accatagaca gaggagaaca ggttctagtt tttattaata 11520
gacgaggatt cgctccactc tatcaatgta gtagttgtgg gtgggtagca gattgtaaat 11580
cttgtgatac aaattttagtc ttccaccagg caagaaatag attaatattgt cataggtgtg 11640
aatctgccta ctctgttaat ttgtcttgtc cggcatgcaa gtctaatagac tttaatatgt 11700
atggagctgg aacagagaga gttgaagaag ttcttaaaag cagctttgta aagactccaa 11760

taattagagt tgatcatgac tcaacaaaaa aagtgggagc tatggaggct atagttaaaa 11820
aaattcattc ctcagacgca gcaatttttag ttggaaactca aatgcttgca aaaggacatg 11880
atthttcctaa agtcacctta agcggttattt taaatgctga taatggcctt ataagcccag 11940
aaattaatgc attagagaaa atatctcaat tgcttattca ggtctctgga agagcaggaa 12000
gaaataataa tcttgcaaaa gttattattc aaacaagata tcctgatgat ataaatctta 12060
ataaaaataa gacaggagat tatatgaaat ttgctttctca atgcctaagt accaatgagc 12120
aaatgaactt acctccattt actactttat gtctgcttag gtgctcatca ccaactcaaa 12180
agagtaatgt agattttccta gagaaagctg ttttaatttt atccaatagg actgatataa 12240
atgttattgg tcccttgcct tcattagttt cgaagtcgaa aggaaattat aggaccaag 12300
tctatatcca tgcaccaaaag aagacttttt taaataaggt attaaagtth ttgacaacag 12360
agtttgaaaa atggccggaa tctaataagg ttaagtgggtc tttcgacatt gatccaatag 12420
acttaagcta aataattaatc ttaattaatt gtccctgggta tattgggttta ttgttttagtt 12480
tattctctgt attaatthttt tctacagtca ccccaaatct tctcgtatt tctgataaga 12540
catccccctt ttgtattttg taagtcacaa agcctggatc aatactcata aaggtatttg 12600

ctttaggttt gtccttaaaa tagttatgta ttccctaggaa aacagatctt gcaatcattc 12660
ttctccctgg cttgcccttt aatctttgtg cgtcttcagg gttgggttata aaccctgact 12720
caaccaatac agaagggtata tcaatagact taagaactct gaaatcagcg tactcaacat 12780
tctttttatg aatttttgtg aaagggtctc ttttaagtgt atccaatctc ttagttccta 12840
aaattttact ttcagaaaatt tttttcttat atattttctgg ataggtttgt cttgcccgcgcat 12900
cctcatcaaa atcaactggc ttttagatttt ttatatcagc ctgtattctt tcccttttgtt 12960
ttttagataa gttccctagca acagtacttg aagcttctatc tgaccatata aaaacagaag 13020
caccctttac ggaagataat ctaaacccat cagcatggat tgaacacaaa atatctgctc 13080
cataattcct agcatcctgg tatctattat ttaaatctaa tgtctcatct ccatctctaa 13140
tcatcaccgg tctgtaccca taagtatctc ttaagggttct ttctaactcc ttcgcaataa 13200
gtaacgtcac atctttttct aaaatattat ttggggccaac cgcaccaggg tatttaccac 13260
cgtgaccggc atctatagca acaacaatat ctctaatact tttattgagg tttttatttc 13320
ttttcacctt aagctctaata tttataattt ctgtatttat agtctgagtt ggggttttgcc 13380
aatggactga ttcatataaa tcaacaacga ttctggtaag acttccatcc tgagatgctc 13440

ttactttttt gattggatag ttgtatggaa catttatctc ggtcggaagg ctcgattgat 13500
taatttccat gacaattcta gagggattttt taaagaataa gaccttacca atgaaaactttt 13560
atctagacta aagttaatac taatttcatt gttaccocatg tcttgtatatt catcgaagaa 13620
aacatcatth ccgctgataa aaaaagatag aaaccaagg attgccaatc tattcataag 13680
tttttaacca ttctttaaag ccattgtcac cagaaaattaa agaaaacctct ctcccttcag 13740
gaagatggct gaaaattatt tttagatcaa aacttctttc atgttgaagc ctttcaggcc 13800
actcaattag cactactttc ttagagtta tttttctact tagatcaaat atatcaatat 13860
cttctgcttc gttagttcta taaagatcaa tatgcaaaaa catataatta ttaaaatcat 13920
attcctcgca gagagtgtaa gttggacttt ttacaagatc ctcccatcca caatttttta 13980
taatagatct agatataaag gttttgcctg ctctaagtc tccctcaaga tgtatttcaa 14040
tttcttgca agaagattta agtatctcca tagctatttt tgaccctagt tggtttgtag 14100
cctcgtcatt tataagagta agtttttttca tcgatttatt aattgtctta ttataggat 14160
taaactagag gcattttaagc ctatttcacc aatatctacc ttaaatattta aacctgcctc 14220
cgaatgaaca gcaacagcaa taatgctgc atttcttata tcaagacctt gggcaagaa 14280

agcagtcaaa acaccagcca gaacgtcacc agttcctcct gaagcaagtt caggtccgcc 14340
gcacgcgcac ataaaagact tgttatttgt tgagtcaaag accaccgttc ccattccttt 14400
taatatgaca atagaagccc cataagtgtc tgcaatttgc ttggcagcag aaatcctatc 14460
tctttgaact tcttcaatag atatgtttag taatatagcg gcctctcctg gatggggcgt 14520
catcaagatt gttttattgg attttttgat taaagatttt gatgatgcca cgatgtgtaa 14580
agcacctgca tctaataataa tttttgaatt atttgcagac tttaatatatt tacccaatat 14640
agttttttgca aaagcagtat tagcaattcc tgggtccatac aatataaacg tatgatcttt 14700
aattttttaa tctatatcga ccccttcaac catatcaaca ccaattgcca ttacctcagg 14760
gtttcttaat aaagatgggc ttacattaga cgtgtctgtg acgagagtta ctagtccaga 14820
gccacaaaat aaacttgcct cagatgcaag tattccggcg ccaccatgc ctggggagcc 14880
agcgcatatc aaaacttttc caaaattacc tttatggcta tcctttgctc tattgggcag 14940
aagatttttt aattctttgaa aggtaaaaact ttgcaatata gacatgaaat tacttaagct 15000
aaattatgtt gaatgctaaa aattcatctt tcagcgaaat aaccaatcct atcggaatgg 15060
aatgatccca taaaaatttt ttcgtcatga ggcaactgcaa tagtgggtag gccaaaaact 15120

gttctagaaa atgaaaaactt atcttctctc ttaagtgttt tttgatttaag tttatgaatg 15180
gagaagggtg gagaacaatt tgttttttct gcacaattac caaaatcatt agcctgtaaa 15240
tctaaagaag ttaaccaaac cgaaccgtct tttaaaaaga tattatctgg actctgaata 15300
aaataactac ctgttttagt atttgatta atatcgtaaa cagataaatt gtcacctga 15360
ttatagttaa cataaagaag tccagagccc tcatctagca aaattccatt aggcccactt 15420
ccatctgttc catcaacttt attaaaaatta ttgtcactcc aaagtactac atgccctgaa 15480
atacttttaa ataatgaatt cattaaccat tctttcatgc tgatatctct tttatacata 15540
tgagatgcat aaaaacttcc atctttttta agggcaacat cattgaaata atattgggtca 15600
gggaccctaa tacagccccg ccatatcata tcccatgaag attcattttt tataatttca 15660
aacatttcaa tcgactcaaa tggcgaatga ttaattacag ctagttgata gaaaccttca 15720
tcattttcaa caagatctat tccgtgggga ttgaatatgt ctagctcgcc tcttatgcat 15780
gaagcgtcac cccaagagct ttctccaaaa gtaatttttg gtaccttttt ttcataggta 15840
tttaaatcca ttaaggcaaa gtaaccaggg gtatgtttctg catatggacc gatccctcca 15900
aattcagaaa taaagaagaa tttattatca ggcgtaatca caatatcttc tgggtttgag 15960

aaattacaaa taaccttgat acggtcatcg gattcacact tactaatatc catttggggg 16020
cctatatagt cagtagatac gacagttaca gaaataaata aataaagaat agaaaaccggg 16080
acggtaatct tataatagtg ctttataaaa atttctaaaa tctttgaagc atgatttgga 16140
agtgcgatca gccaaacccc ttttaaaaaa gatagagccc ccataacaac aaaaattact 16200
gaccatagcc tgtcagacca ttccgggctga agaagaccag ttaaaaaaag aatcatccca 16260
aagaataaag ctaaatagcc tgagattttg actttcgatc caacaaatctt cgaaaagaga 16320
tatgtataaa gaggctttat aagaactagt aaaccacagg ccaagaaaaa gaatgctaag 16380
tagtaattca taagttagtt tttatataaa tgctccttaa taatactaac aagtttctaag 16440
ggcttgtcca atggaacatg gtgagcagct ccaggaaccc cctcaaaaagt cataatgtca 16500
ccatatgtat ttttaatat gtccaagata ctccggagg ttaataagct gtcttcaccg 16560
tggatgaaca aagcagggca gccaaatgaa aatgtgtaac cgaataaacct ttcaagactg 16620
ctaaacatga catcatcaaa tttccatctc caccagcct caatatTTTT tactgagtgc 16680
tcagcaatgt atctttaagta ccaatcattc gtacaaatctt gcttaggcac taacctaac 16740
cttttaataa tatctgtctt gtcttgatag tgcttgatca ttctgagagg agaagagtgt 16800

tgattcgggt cataatccgg tgggtcttata aatgtatcaa taataataat attattttatt 16860
agatcccttc ttccagatgc aacgtaacca gcaacatgtc cgccgagggg gtgtccaaca 16920
ataaaaaatat ttgaaaatat ttttttatcc ttttcctttt caattacaga gacaatacat 16980
tctccaaaat ctttaatgcc atatgaatct ctaaaagaag agtcacccat gccagggaaga 17040
tctattgcaa ctatatttgc gcagtctcta aagtggggcg caataggatc ccaccatttt 17100
ttatgagcac ctgttccgtg aataagaatt attaaatctt tgctttcatc tttggagttc 17160
cagctagaat aggatatatc cccatgagga ttcttgataa tctctgagct aggccttgctc 17220
tcaatggcat ctttgaacca ctgtggggca tgaataatgt cttgatttag attgttagtt 17280
atttccataa acagtattct aagctataaa aaataaaaat atgaataaac ttaatttaac 17340
gccagcagca actgttttag tcctaaagga ttctcctgat gggatggaag ttttgatggt 17400
aaaaagatca agtaggcctc ccttcggaga cctttttgtt tccccggcg gcaagattga 17460
cgaagggtgat ttcaataata agatagaaga tttttgtgag ggcgtgactg ataaagaggc 17520
ctccataaat cttggattag atttctggagg tctagcatat tgggttgcat gtattagaga 17580
atgctttgag gaggttgga ttttacttgc taaaaaaaag agtggggaag atcttgatct 17640

agatggagtc gataaacata aatatcaaaa atatagagag atgttggttaa ataatagaat 17700
tgatttatat aaaatctgtt tagaagaaaa ttaattcta atgcctcaac aaatagcccc 17760
tttctcgcat tgataaacc ctgaataga aactaggaga ttgatacac gtttttttat 17820
tgcccacctc ccaagcatc agaccggaga acatgatggt agtgagctca tagacagtgt 17880
ttggatttca ccaaagaag cgctcaaaaa atctcgttcg ggtgagatgc ctatgattat 17940
gcctacaata aaaaatttgg aacaatgtgc acaatttgat tcgggctcta agctttttaga 18000
aaatcagagg aatctctcaa atgaggatat ccaccaatc ctgccaaaagt tttttaaaga 18060
agatggtgag tggaggggtc tattgcctgg agataaaggg tatgaggatc attaaataat 18120
atggacttaa ttactaaat aacagctccc aaccctggtg ttttcaactgg gggtagggact 18180
aatacttatt tgattggcaa agatgatata acccttgctg accctgggtcc aaatatatct 18240
gagcatctag atgaaattat caaagcaggg gatgggaaaa taaaaagaat ctttgttact 18300
catacccata cagatcatc cccagccgca ttgcctttat caaaaactct taatgtttcca 18360
atgtacggaa ggctagtaga tgggtgaatcc tcatgggagg atgaaacatt tatcccagat 18420
attattttaa atgataaaga tattattgag acagacgaat atacgttaga agtaatacat 18480

actcctgggc acgcatctaa tcatttatgc tttttaataa aagatacgaa atgccttcta 18540
acaggcgatc acattatgga cgggtctacg gttgttattg ggccaccaga tggcaatatg 18600
acaagctata tcaattcatt agaaaagtta ctagattttg atattgattg ctttgcgcct 18660
gggcatggaa attatatcca tgagcctgag aaaaccattc aatcaattat taggcacaga 18720
ctaacaagag aaagaaaagc tcttagaaaag ctaggagagg caggaatctc atcattagat 18780
aaacttacta agcttgttta tgatgatgtg tcagagatgc tccatcctat agctaaatat 18840
agtctagaag cacatttatt aaagcttata gatgaaaaga aagttaaatt agataaaagat 18900
ctattcgaaa taattttaac ctttttattt ttatgtaaga ctttctcctc aatagcttct 18960
atatcaatat catcaattga gtcttcgtta ttatcaggta tctttttttac atcttttttca 19020
attttaaggt cgataggaga ggctccaaga tcaaaaagta gctcccttac attttttgaa 19080
atagtatcct caacacagtc atcttgatcg tatgcctcac gggctctctcc ttatcatta 19140
atgggaaaagg gtctttgggg agggccattt tgcatgcatt taatagtagc aactggtgag 19200
tagtaatcgt cactagaata taacttatca agttcttttg gtgatatgga gcaaccaata 19260
attccaaata gaataaatgg cgctaagagt cttttcataa actttgtttg ttttctataa 19320

gagactcaac aaccgaagga tcggcaaggg ttgtcgtgtc tcctaaatta gatagatcat 19380
tctcagcaat ctttctaaga attctgcgca taatttttcc cgatcttgtt ttaggcaagc 19440
ctggagcatt ttgaattaaa tctggttttg caatagctcc aatttcttta gcaacaaatt 19500
gtttcaattc ataactaaag ttgtcatcaa atgattcatt tatcattaaa gtaacaaaag 19560
catatatcc ttgccccctta attggatgat caaaaccaac aacagctgct tcagcaattt 19620
taggggtggag cacaagagca ctttcaattt cagctgtacc tagtctgtgg ccagaaacat 19680
taagaacatc atcaactctt ccggttatcc agaagtatcc atcctcgtct cgcctggctc 19740
catcaccagt aaagtaaata tctttataca taccaaaata ggtgtcgatc attctttggg 19800
gatcaccata aatacttcta atttgactag gccaaagattg ctcaataact aaattacctg 19860
cattagagcc ttctagcgta ttcccatgct cgtcatagag agatgggctta actccgaaga 19920
agggcagagt tgctgaccca ggttttgttg gagtaatacc cgctattgga gagataagta 19980
cagaaccagt ttcggtttgc caccaagtat caataacctc gcaattagat ttaccgacaa 20040
cactgtagta ccaatccac gcctctggat taattgggctc tccaactgta cctaaaattc 20100
ttaggctatc tctttttgtc tttttaacag gatcatcgcc ttgggccatc agagctctaa 20160

tggcagttgg agctgtataa aaaatactaa tgtcatgctt atcgcatatc tcccaacacc 20220
ttgatgctga agggtaggtt ggcactcctt caaacataaag tgttggtgct ccatttgaaa 20280
gaggtcgcta caagatataa gtatgtcctg ttatccatcc cacatctgca gtacaccagt 20340
atttgtctc tggccttatt ccaaaaagat attgaaact aatatgagcg cctaatagat 20400
aacctgcagt agtatgtaga acacccttg gcttgccctgt agagcctgat gtatagagaa 20460
tgaaaagagg gtcttcggaa tccataggct ctggagcaca cttattagaa acatctttaa 20520
caagatcttc ataccaaaca tcttttttat catcccaatt aatttcgcca ccagttcttt 20580
ttataaccag tgtattttta acatctggac agcccagaag agcctcatct acattagatt 20640
taagtggcac ttttttgccg ccccttaaac cttcatcagc agttataaca attttacaat 20700
cagcatcaag aattcttatct ttgagtgatt ctggagaaaa gccaccaaaag acaacagagt 20760
gcacggcacc tattcgctg caagcaagca ttgcgaacgc agtctcaatg atcataggca 20820
tataaataca aactctcgag cccttttgaa caccaggctc ttttaaaaaa ttagcgaact 20880
tacatacttc gtcatggagc tctttgtagg tcaattcttt agaatcagca gggtcattctc 20940
cttcccatat taatgctatt ttgtttggat catttttctaa atgcctatcg atgcagttta 21000

agctaataatt tgtttttacca ccctcaaacc acttcgcatt attaaattga ttattgaatg 21060
ttgtttttgaa gtctttccatc cagcttatgt tttcgttttgc tagattttta aaaaattttag 21120
aaggatcttc tatggattgc ttgtaaaagt ctttgtattc atcaaagtct tttatataaag 21180
gattacttga atgtttttggg ctataaaagt ttcgaggcat tcttaataa ttgaaggttg 21240
ggggttaatg aaattctttc ctttgggatt ggacattatt tttgtaatga gcgattcgtg 21300
atcgctatca ttggttttcta aattaatc tgcagcatta attattaata aaggggcaga 21360
actatagtct aagaaaaacc ttgagtatgc atcattcagt ctttccagggt agtcaagagt 21420
tagatatgtg tcgtttaatat ttcctctctt agtaatcctg tcttttaaca catcaatagg 21480
tgcctgaaga tagattacta ggtcgggtgt tggcgcgtcc agggtttagat ggtcatatac 21540
tttgtcatat agatccattt cctcgttaga aagagtaacc tcagcgaata atcgatcttt 21600
ttctattaaa aaatcagcaa ccctcactgt ttcaaaaagg cttctttgtt taagatcttg 21660
aatttgttgc attctttgaa acaagaagaa aagctgagt gctagagctg attggcttgg 21720
gtttttataa aaattcttta agaagggatt ctctgccggt tgttctaaaa aagaatcata 21780
attaaaatgt tcggctatct tatttgctaa agtagttttc ccaacaccta tcgggtccttc 21840

aattgcgata tatttttgaa gtggtacttc tttaatagct gggttcattt ataacgtata 21900
gtttttatcc tagattatct caactcgtta tagaagccaa atatttttgc agattctttca 21960
tcactttttt cattcatctc agcaataatg gtctctctca atgagtcato attttttttg 22020
aaatcagtc accacttccc aataccactc ttgtctccag aagctctctc tctacaagg 22080
agaaaaatcat atgcagctct aaaccttgga tgtcgaagag ttttataatgg ctggctacca 22140
attctgctat gaagtthttaa ttgaagtacc caaatatcct taatatagct tgaaaaatttt 22200
cttggtattg ctgtgatttt ttgttggttcg cgaagtacac catccataga tcggaaaaaat 22260
ttctcacat taattttctc attcttagaa cacttttttca ataacgaggg ccataacaat 22320
gcagccataa gaaagcctgg tgtaattgac tgttggtttt taacctatc atcagtattt 22380
cttagagcat gtgtcataac attgctggca aagtcattcc tacttgatc agaaaaggata 22440
agataatttat ttaaatgaaa tgagcagagt ttttcaaat tcttttcacc cattccattc 22500
aaaaatatth tgcagaactc atcaaacat cttagcgttg agatgcctga taaaagatgg 22560
cctttatcat agatggcatc cttaaccaga ttatctatct taaaattgag tttgttacta 22620
aatcttatag ctcttaagct tctcactgga tcttcttga atcgtctttg gggatctcca 22680

atagatacaa taaccttctt gtgtatatgc ttaagcccat cgttatgac ttctattttt 22740
tttghtaacag gacagtagta aagggcattc acagtaaaat ctcttctatg acaatcttgt 22800
tcaagagtgc cccaattatt atctctaaga atctttcctg tcgaatcagt aacaataatt 22860
tcaccatctt cctgatcact tcctgatctg aaagttgcta cttcaagcaa ttcactccta 22920
ttaaaaacat gaaccaattt aaatcttttg ccaattattc ttgaagcctt aaatgttttc 22980
ctgatctgct caggcgtggc attagtggct atatcgaaat cttttggctc tagtcccgtt 23040
agtgcacgc gaacgcaccc accaaccaga taggcctgaa aattattttt ttgaaggctc 23100
tgaacgacag atatagcaaa ttacttatac ttattattat ctatcaaatc taattatgaa 23160
tttttatgat ttaagggcat tagccattta gctgcttttc tttaatttcg tcaagtgttt 23220
tgcaatgaat gcaatgagtt gcagtggtc tagcttcaag cctctaatt ccaatttcat 23280
caccacagga ttcgcaccaa ccataatcat cttgcttaat ttgttcaata gatagaccta 23340
tttttgctgat aagttttctc tctctatctc tgggttcttaa ttcaaaagca aactcttcct 23400
cttgagaagc cctgtctact gggctctgcat aggtttcacc tttagctcta agatgatcaa 23460
aagttttttg catttcatcc ttaagatggt ctttccaaag aagaagaacg gcaacaaaat 23520

gtttcttcat tgctgcactc atatatTTTT cacccttctt agatttataa ggtgcaattt 23580
tagattttatt attagctatt gttgcttttg cagacttctt ggctgcaact tttttaactg 23640
gagctttttt ggtaacagtt tttttaactg tagctttttt ggcaacagtt tttttaactg 23700
gagctttttt tactgtcttg gattttttt cgaccatgta agattttata gaattttagg 23760
gtggagaaaa taccagatac tgacaaaaatt agctagtcat tttttaattt atttaatact 23820
ttcaagtacc catccgagct aagtcttggc ccaatgtct caacaacctt ggaagatgca 23880
tagtttgcaa acttagcaca tgcttcaata ttattccctt gaaggtaggc atgcataaac 23940
gatccggcaa acatatcacc agctccattg gtatctattg gagttatttc ttctgcttga 24000
gcatgcttct caaccctttt atctataaca acacttccat cggcgctttt tgtaatagca 24060
gtcatatagg gcttttcttt ataaaagcta acagcatcat caaggctttc ttaccagaa 24120
aaagcaacag cttcatcatc attacagaag atcatatcta ttccatatga ctctattaaa 24180
tcaaattttt ctttaaaacc atgaacaata cctgcatcag aaagagacaa ggcttttcttt 24240
acgtcccttg ctttaagggtg ctctaagact gaaataacag cattaaagtt atcgtcactt 24300
gttaccatgt agccttcgat ataaaaaatt tttgaatttt ctacaacatc aaaatctata 24360

tctgatttac caagatacgc actaactcca agcatgctgc tcatagttct cttagcgtca 24420
ggagtaacta aaattaagca tttcccagtt ggttgatctg tattttcaga gctgacacca 24480
atatgtttga ctccagccga cctgagacta tcaagatagt ttcttccatc ttcatcatca 24540
gaaactctgc atacatgatg gcaattcgaa ccataatttg ctgcagcaac aagagaattg 24600
gttgagagc caccgcaatc agaaatcgat tcggctccca ttccaataag ttgctaatt 24660
atagggtgcct gttcttcaga agatgaaaga gtcataagat cggctacaag gcctacactt 24720
gataaaaaatt catggctgac tttatatatga gtatctacta aagcattttcc aagggcgcta 24780
atatcatatt tcatgtgtta ttccctttttg tattatttgt tttaactgttt ctaacgttct 24840
attaatttct tgatcccttat gcattgctga aatgaagcct gcttcataatt tagagggggc 24900
aaagtatatatt ccacttcttaa tacatgaatt taaaaaattt gaaaatagta catcatcagt 24960
ttttgcaaca tcattaatat tattaggag ttcttctgaa aaaaaaatc cgaacattcc 25020
accaattctg tttatagaaa acggaatacc tgactcaatc attaagggttt tcatcccatc 25080
caaaagcaca gatgcatttc tttctaactc ttttaaacgga ttttcttttaa tcaacaattg 25140
caacaaagca gttccctccag ccatagctag tggattgcca gacaaagttc ctgcttgata 25200

gacaggacca gaaggagcta gatagttcat aattttcttct ttgcctccaa aagctccaac 25260
aggaagaccg ccacctatta ctttccccaa agcagttaag tcaggagtaa tattataaat 25320
ttcttgagct cgcctagcg aaactctaaa gccgctcatg acctcatcaa atattaaaaat 25380
agaattattg gctgaggtagg ttccctctcaa taactttaag aaatcttcat gacctggaac 25440
aaagcccata ttccctgcta ctggttcgac tattacggcc gctaagtcac cttttatctc 25500
attaatatatt tctaaaaact gttcttttatt attgtattcg caactaaatg tataattttgc 25560
caaatctgca ggaacccttg gagagtcagg taagccaaag gtggcaaccc cagaaccgcg 25620
cttaattaaa agagagtcaa catgaccgtg ataacagcca tcaaatttaa taattttatt 25680
ccttcctgta aaacctctgg ccaatctaat cgttgtcatg gttgcttcag tacctgaatt 25740
aaccattctt attttttcaa ttgaaggaaat gcatttttta attagcctgg ccacatcaga 25800
ttcaaagactt gtcggggcgc cataactagt tccaagcgca acttgatttt taattgcacc 25860
tacaatgtct ggggtgtgaat gaccatttat catcggacc caagacccaa tataatcaat 25920
atattcatta tgatcagcat cataaaggta tgcgcgggat gctctttcaa aaaatatagg 25980
attgccattg atatttttga atgcccttac tgggtgaattt acccctccag gcataagagt 26040

cttggcctct ttgaataagg ctatggattt atcaatttta ttatgtgtca aagctaattc 26100
ttcttaattt ttaatttcga tatgatatca acttaattct tattatgtat ttaattgttt 26160
tgaaaaattca tccacagtat tccagttggt gaactcataa gtatttgtaa catctgtagg 26220
accctttgta atccacataa tcaatctaata catgtgttta tctataaatt tataatttgg 26280
ataatctatt ttgcctgcaa agactgctaa tttcattggg ttccaaggag acagctctaa 26340
aaatttttgc atatacgggt ttgttttcagg tgtatttttt tcaggctttc tcgcaactac 26400
attaactgaa aagaaggcat tttcttttgt ttcaagacaa gcaacatttt tttgaataaa 26460
ttcataaagt tctggtttgt gtttgccata cctaagtctc gcaccaataa taattttatc 26520
aaattgatat aaatctaact ctactgcttt tgcaatatgt attatttttg aagattcaga 26580
cacatctaata ttagagaaaa tttttttaca aatttcaga gtttgcccat cggttgtaga 26640
gtagattagt agagttgatt tcataaaatat attttatatt gttaattaaa aaatttcattg 26700
gcacgcgcaa aatataaaca aaaatatagt atatttaaaa aatgattaaa tatattatta 26760
ttatcgcttt attcatcagc aactttgcat atgcaaatat tgatgcagct aggtgcgcag 26820
ggattagttc agatcaagaa agacttgatt gctatgattt aatttttaa gctaattgatg 26880

agctaccttt agatagtaat attaaaactc tcatcacacc agctatcaag gctgtgaccc 26940
cagctgattc aataaaaaatt gaaaataaaag caacaaaaaga aaaggatttc gggcttccaa 27000
aaacaaaaat caaaaactct gcaaagaact caataaaaac ctcggtagta aggattaaaa 27060
aaacaaaaag tggtaaatta attttactt tagaaaatga gcaagaatgg actgctgaaa 27120
cttcttatag agcaaggaat atgtttaaac cagaaaccgc agtcatttta gaagaggccc 27180
tggttagtgg tttttatatg attaatataa gtaataaaca gaaaattaga ataaagaggt 27240
tgaaatagca atgaccatac agagtatagc aacgaccgag ggtgccatga aaagaataag 27300
atctgttctg gggttcccaag atggctcaag cttcagggtt tatgttaccg gaggtgggtg 27360
ctctggggtt caatatggct tcaagtttga taacgatatt gcattcgatg atgatgttat 27420
aaattgtgga gatttctcgc ttttaataga ctctatgtcg taccctatc tttatggatc 27480
aactctggat tttgttgaag atctctcagg ggctaaatc gttattaaaa acccaaatgc 27540
caaaacaaca tgcgggtgtg gagagtcatt tacagtttag attttgtaat tgaacccaaa 27600
agtcctttta cattcttgct tgatgtctga actatcagag cctcattact gactctttta 27660
tatcccatcc aggcaaatgc catggactca atagcaaaaa catcatggcc taaatcactt 27720

gagaggacga tatcattact agccatctca gaaattcttt ttaccagata cttgttatgg 27780
gcaccaccgc cacaaatgac aatgtcacaa ttattatgac catTTTTtatg gattgaattt 27840
attatagatt ttgctgaaaa ttcaaccagg gtacatagaa tatcttcagc ttcttttttt 27900
aagaatttct tggataatat tttatatata aagagctctt tgcctgttga ttttggacat 27960
tttcttttaa aaaaattatt ttgaagcaat cttcctaact caatgtgac tacttttctc 28020
ttggctgcaa ttgcaccatt tttatcaaaa ggaatctgaa gaaaatcact acaatatgca 28080
tctaaaaattg cattaccagg gcctacatca gttccccata tgtcatttct atttttttaca 28140
aatgagtaat ttgaaatccc ccctatatat aaaataattc gtgggtttct agcttttataa 28200
aataattgat tatggaattc cggaacaaga gggggcgcct cgccaccag agcaatatgc 28260
atatattctaa aatcacttac caccagaagg cccgtttctt ttgcaacaat atttggatca 28320
ccaatttgca tagaaaatgg aaatctctta ttaatttcat gcctaattggt ttgtccagaa 28380
atagcaacac actctataga tgattttctta attttactaa atccaatcgc ctcatataa 28440
gatttttgaaa ataagaatcc tatctcttta ttgatatgac ctaaactctga caacgagctt 28500
tcgttatttt caataagttt ttttactttt aatcttaggg attttgggaa tttaattgaa 28560

tggaataact caagatgtat ttttgtacca atgctaagaa atgaaatatac aatagcgtca 28620
tgactagttc ctgtcattgc tccaatatag atcttttttac tcattactac ttggagaaaag 28680
tttattgaac tcctccattg ataatttatt attcttttaac agtgtatcaa aacttgcacct 28740
taaatTTTTT ggtactgggt cagctgatgg taatttttact tttataggat ctattctttt 28800
cttacctatc ttaaaactcat aatgtaaatg aggccctgtt gccagtcctg aactcccaac 28860
gaaaccaata gtatcgccct gaaaaacttt tttgcctttt cttataccct tactaaaattt 28920
ttctaaatga caatatctag tcgaatactc atttgtatga tttataacta tctcattacc 28980
acatccattc ctttgaccag aaaaagaac aatcccatcg cctgtagttc tcaccggcga 29040
tcctcttttt gctgcataat cgactccatt gtgagctctt attgtatgaa gaacaggatg 29100
cattctgttt ggattaaaat gagaactaat atatgcaaaa tctaaaaggg ctcttaaaaa 29160
ggcctttttg atattatttc cattttcatc aaagtactgt tttttactgg cttcagtga 29220
aaacctattt gaaaaatagg tattaccatt gttaataaac ttagcaatta cgatatcgcc 29280
attttttact ttctctccat cactataagg agtgtcataa attacatgaa attcgtcacc 29340
ctctctaata tcaaaaaacga agtctacatc ccagccaaag atataagcaa agtccataat 29400

caacttttcg ggaatatttg catctagagc tgccctcataa aatgaggatt gaataatccc 29460
gctatttgtaa gattcaatta actcaatact tttgcttata ttcttaaata ctattttctgg 29520
agtaaggcta attgaaatag aatttatattg atcttttcattg atctcaattc tggtttagctc 29580
ctcaccagaa tattcaaaga gcattttttc ccctggttta atgttagcaa tgatatatttt 29640
ggagtctaact ctaaaaattt tataagcagt atttaaaggc actgaaaagt ttccaataat 29700
tattgaaaga tttttctccat ctgacacctc atgcattctga tatgtttttcg tttgctcgag 29760
aggtatagag aattttttcag taatctcaat ttcttcaact ggcaaagatt catatgttttc 29820
tatatcaata taaagcatta atattaagac aatagaaaca gcaaaaagaaa caagtaccgc 29880
tcttttttga acttttttaa aacctatcat tagtcctctt gaatatattaa gagttcagtg 29940
ttgccaccaa atgctaccga gttctttgag acaacctttt cgtaaagaaa ttgcagaagg 30000
taattaggcc caccagcctt ataaccgctg ccggatagat tttgtccacc aaaaggctga 30060
gaaccaacaa cagctccccc catatctcta tttatgtaaa tatttccaac attacacttg 30120
tcactaaata tgtcggccct ttttttctact ctagtatgaa taccatttgt aaggccgaag 30180
ccactatcat taatatatttc aattagtga tccatttcat ttgattttaa tctaacgata 30240

tgaagtattg gtccaaattg ctcgctctttt agatcagaaa tattatcaat ctcaattatt 30300
gttggagcta caaaattttc atctacatta tcagcagacc taaaaataga ataatttctta 30360
tcttcaaaac ctttgacgta agcattgaga ctatcaagag atgttttgct aattattgga 30420
ccgatatctg tgtcaagatt ttgagggttt ccaattttta gttctttcat accaccctta 30480
atcattgata aaagatcgtc atatatctcg tcttgtaac agagaactct caaggcagag 30540
catcgctgtc cagagctatc aaatgctgaa cgaataatat catcgggtgc ttgctcgaga 30600
agtgcgcttg aatcaacaat cattgaattt atgcctcctg tctctgcaat caaaggaata 30660
attgattcat ggttacttgc aaggctgctt tgaatttttt ttgctgtttt taaagaccgg 30720
gtgaatgcaa cccctttgat gttattaacc tttgaaagca tgtcaccatg aatgccgtct 30780
cctaggatta aatttaaagc atcttttggg accccaatct catgaaaatt gtccacaatg 30840
atgtagccaa gaattgaagt atgttctgag ggcttaactg ttactttatt tccacatgca 30900
agggctgcac ttatttgccc tataagtatt gcaacaggaa agttccatgg actaatacat 30960
aaaatatggc ctttttggtga gtaagatagg gcattgatct cgccggtggg accctcaaga 31020
atatgggtctt ctgttttgag gccaacagct tgttttgcat agtatctgag aaaatctata 31080

gcttctctaa tttcatcaat agtattttgt actgttttttc cggcttcatt cataaggtaa 31140
taaattagct cagaaggatt tgcttcaatg tcgtctgcaa ttttttctaa aatggatgct 31200
cttttttcaa catgcatcaa agaccattca ctaataattt gttcttcaag gctgcccttt 31260
atataatcta gatcatcata tgaggctgta ccaataattc ttccatctgc aagagatgag 31320
atgtcatgag tattgggttt cttataatcc ttgctttat aaatcgatga agcctttatt 31380
tcttttgaat caaactttcc aagctcctct tcaagcattt ctaaattcac ctttccactt 31440
aaatcaaadc ctttagagtt cggctctgtca ttaaatatat ttcttggcat tggatatctct 31500
tttttttcat cctcaatttt taaatggggc cctctggcaa gccaaagcaga atctgttttca 31560
ggatcaagta acctattaat aaatgagctg ttgctccat tttctaatag tcttctttaca 31620
agatagggtg gcaagtcttt atatttgcca atcggcgcat aaattgaggt attttttttca 31680
gtatttaaga tttgtttgc ggacttataa agcagctctc ccattccaaa tagtctttga 31740
aattcataat ccttatttga gccaaagatga tggattgcag aaatgggatg tgcattatgt 31800
gtagcaaat tagggtaaat tttttccaca ttaaaaaatc tttttgcaca ggctaaataa 31860
gctaagtctg taacagattt ttttgtatag acaggataac catcatagcc atagatttga 31920

gcatgcttaa tttcataatc ccaatatgct ccttttacga gccgaacatg cataggtgct 31980
ctgttttcta gtagctcttc taaccaatct attgtggcta tagctctttt gccatatgcc 32040
tgaacagcaa taccaaatcc tttccagttt ttaatatattg gtgaaaagggc catctcttta 32100
ataatctcta aactaacggc cagtctgtct tgctcctcag catcgattgt aatttctaca 32160
tcttttagact ttgcatactc tgtgagctga ataagttttg gaagtagatc agacttttatg 32220
tcttttagct ttttcatttc atatctagga gataaagctg atatttttat tgagacacca 32280
ttaattgtat tttttgtcag atttatcttg ccgacttcat ctatagcatt cttataagac 32340
tgataataag tgtcggcttg ttcagcgttt ctgctgcct ctccaagcat gtcaaatgaa 32400
taaatttcat tttctatatt cttaattttt tttatgtcat caaatctct gcccataaca 32460
aattcttgac tgagaatatg catggcacca actacagcat ttctaattgg aaactcgcca 32520
gatttttgaga ttaaagaact taatagggcg cttggggttt tggtccactc atcgggtgta 32580
gaaactacct tgcctgcaag aagaagaccc catgttgatg cattaacgaa aacactatcc 32640
gctttgttta aatgctctat ccacgcaccc tcagataaact ttcagatat tattagatcc 32700
cttgtttttt tgtctgggat tcttagtatt gattcggcta aacacatcag agcaaacct 32760

tctttattat ctagcccata ttcgcttaga aatgcatcta gttttgttct ctcgctctta 32820
ttctctctgc aagcatcaat tattttattg agcatttttt gaaatcgaag ggtcatttaa 32880
aaaatcagaa ttactaataa gatctgaaac tatttcttgt tcaggataaa actgtttact 32940
tgttaatgtc atagctatta ttttaatctt atagtccaa gtatcaactt actgattcta 33000
ttaatgactt aaaaacatta tgatagtcac atgagtgatg cacttaaatt aattaaaacga 33060
ggaaccgacg agatccctcac agagtctgat ttaaaaaaga aattagattc tggaaaaacag 33120
ctaatcatta aggcggggtt tgatccaaca gctccagacc ttcatattggg tcataccgta 33180
ttattaaata agctaagaca ttttcaagat ctggggcata aagtaatttt tctaattggt 33240
gattttactg gtcaaatgg agatccctcc ggtaaaaaata aaactaggcc aacacttact 33300
tcagaaagaat taatttcaaa tgcaaaaaaca tatgagaaac aagtttttaa aattcttaaa 33360
aaagaattaa cagaagttaa atttaattct gagtgggtgca acaagcttgg tgcagatggg 33420
ttgattggtc ttgcatcaaa atataatgtt gcaagaatgc tggagcgtga tgatttttaat 33480
aaacgttata gcgcaaatca aagcatagct attcatgaat ttttatacc ccttgttcaa 33540
ggatatgact ctgtagccct agaggctgat gtcgaatgcg gaggaacaga tcaaaaattt 33600

aatttgctag tagggagaga gctacaaaga tcctatgggtc aagaacctca ggttgtttta 33660
actgtacca ttctagaagg cttggacgga ataaataaaa tgtccaaatc attaaataac 33720
tttatagcaa tagatgaaga gcctaatgat atgttcggta aaataatgtc tatttcagac 33780
gagttaatgt ggagatgggtt tgagttactc agctttacct cagagaaaga aataggagtt 33840
cttaagaaga aaatggaaga agggaccaac ccaagagata ttaagtttct tttagcagaa 33900
gagttagtag atagattcca ttcagaggggt gatggttcga aatgcaagga agctttttctt 33960
caaagatttc aaaaaggtca aatgcctgat gacattcctt ccatgtcagt tgatgttggg 34020
gctgagggca ttccattagt aaacctttta aagaattgtg agatgacatc aagtacatct 34080
gaggcgatga gacttggtta acaaggcgga gttaaaatcg actctgtaaa aatagaagat 34140
cctaaaaatgc taatttcaa aggccaggag tctatttata tagtccaaat atttcttctt ctttttatag 34200
ttaaaaaatta aaacataatg aaaaaataaat tagtccaaat atttcttctt ctttttatag 34260
ttgcgtgtaa tcaggatagc cctaataatta aacaaattag taatatgcag tatttttattg 34320
ataatgaaat aagggaagga atctattctg ttgaaccagg cttgcaatac tcaattatcc 34380
aaaatggaga tcaagttctt gaatcaccat tgctgcagga tacaattaca gctcattttc 34440

acgggaccct cactgatggc tcagtttttt ggagtttctgt tgaaatgggt gagcccttaa 34500
cagtcgaaact atcaggtcta atagttgggt gccaaaaaat aatctctatg atgaagaaag 34560
gtgatgaatg gagagtatat atcgacccaa gtagggccta tggcgatgag ggccggcctg 34620
ggatacctc aaactcaatt ctatatttttg atattgagtt attagatata caaaaaaact 34680
aaccttatac tataacggat agagcgtatc cataaacctc agcaaacctg ttaataattt 34740
tatttttagg tgttccggct ccatgaccag ctcttccttc aattctaatt aatatggat 34800
tatcacacc ctgggactct tgaagttttg ctgcaaaactt gaatgaatga gaggtacta 34860
ctctatcac tcttttagct gtagtaatca gagtagttgg atagcactca ctttcaacaa 34920
tattatgcag aggcgaataa gctaataagt tttcaaactc atcttttctta tctggagagc 34980
catagtcact ttcccaggcc cagcctatag taaatttatg aaatctaagc atgtccaaga 35040
cgccaacttg tggaattgcg actttaataa aattaggatt ttgcaacatg gtcgcagcca 35100
ctagcaaaacc tccattagag ccaccttgaa tagcgggtga tgatggagaa ccaatttctt 35160
gtgcgtgtaa aaatttagca gaatatgcaa agtcatcaaa aacattttgt ttattaaaaa 35220
gcctgcctgc atcatgccag ttatcaccat attcaccacc acctcttaga ttaactacag 35280

caactattcc cccttggttc atccaagtga ggtagctttt actaaagcca ggtagtcttg 35340
agataattaa gccaccatat ccataataaa gtattggagt attactgtct atttttaagg 35400
actttttgta actaagatgg atgggtattt gtgttccatc ttgggatgga aagaatttaa 35460
agtcagatgt aaataaagtt gagtcgtgtc cttaagatc ttctttccaa aaaagtctctt 35520
gtgacataatc agttagatta attttgtata tttctcgagg agttacaaaaa tttgtaaatg 35580
aaaaataaga tacctcatct tcaatttcac caccaaaacc acccattggtt ctttttctttt 35640
ctgttgctag cttattttta tatgctcctt ttaaatcaaa gaagtgaacc tcagtaaaag 35700
tatcaacgag ataagaaact acaatagaat tatttataaa gctaacacta ctaatagaat 35760
tagtactttc accaacgact tcattccaaa caaatgagcc attttttatt gtaagtata 35820
ctacttttcc gttcgcagca ttttcagttg agtaaaaacca aaaagtatca ttcttgctttt 35880
ctaaaaagct ataagcgcct attagctcat ctattaaagg aataaaaggc tgatcagggc 35940
ttagctgaac atagagtcta tttcttttcat ctgtgccttc accaatagat agaaaacttaa 36000
ttttagaatc ttttacaaca cttattcccc aactccatct tggctttttca ggatttctcat 36060
aaacaatcac atcctcatct tgagcagtac caattttatg aaacattaac tttggagctg 36120

tattaatatc ttttaaaagc tcttcgatg gctcgtcgtgta tttttggtaa taaaatccag 36180
aatcatcatt ttcccatgaa gcaccagaaa atttagccca ttcaatccta tcatcaagag 36240
tttttcctga ttcaatatca agtactttcc aggttctcca atcagaccca ccatctgata 36300
ttgagaaaagc tagaagagat gcatcattac taacgctggt acttgcaaga gagatagttc 36360
catcttctga gaactgatth ggatcaagta aaactctatc ttggcatttct tcacagtcct 36420
taatcatcag ctgcttttgt tgccatgacc catcattgaa ataataaaaa gtttttttgt 36480
taacctgata aggcatgctt atcgaatctg tatcccaaac ctcatctaaa ttcttagcaa 36540
tagatttttt atatttattt tggcctataa attttttgtt gaaattattt tgtctctcga 36600
cccagtctgt tgagtcctca ctgtaaaaat cctccatcca tctataagca tcttcgatta 36660
ggtatccatg gacttcttca ctaaaaggaa ctttatttga ttcagggtat tcaaaaattat 36720
ctgttttgatt tgagcaacta actagaatta atagagctgc catgtatatg tattttattgc 36780
gcataaaaag cataatactt tacaattacc ttatgaatca attcaatcga gcatggcaac 36840
ttttacgcaa aaaattggaa gtttttttatt gtacttgccg caccatcat gatgcttgaa 36900
atagcaatgg catcaatggt aacccaata caaaacgtta ctcagcctga ggatatctta 36960

gaatthttca atgagaatat agccttttcta ggctctgtta gtctcttagg cgtagtttta 37020
agtatggctt ttatgggagc gcttttttggtt tcgtatgcat ccatagaatc agaaaaatgag 37080
attgagcctc taaatgcctt attttttaggc ataagaaaaat tctttccact tctggggggct 37140
tatcttatag cttcagttgg tgtctttttt ggtatcttat tattaatact cccagctttt 37200
tatgtagcag caaggctttg catttttctt gcatttatta tgcttgaaga caaaggagct 37260
atagaatctc ttaaattatc ttgggaaaaag acggatgagc atggcaccac tttgttttggg 37320
cttaccatta ccttcttttc ttttaacaatg atttttgcat cagttgcccataccattata 37380
agtcgggat taatgcaatt agttgtttctt gcaattattg aatatgtaat agtaattcca 37440
tggggctatg tataatttttag tttatacaag tcattaaaaa gatattaggc agattaaacta 37500
gaataaatta atacgaagct tttcgttttac aaaaaaacac accactctta taatccttcc 37560
tccgaagatt tgcactttgt taattcatgc aaacaaactt gggctctgtag ctacagcttgg 37620
ttagagcgca ccctgataa gggtgaggtc ggtggttcga gtccaccag accaccatt 37680
ctttgaacat taacttttat ttcctttata cttatttcta actaaataag gatagtatta 37740
tggctaagg tttagacaag caaaaaaacg acaagaagaa aggcaaaact ctaaaagaaa 37800

aaagagcggc aaaaaaagaa aaaaaagaat agttttttgt agaagactat tattatctag 37860
tcgaccctagg gccatcttta aaatctatatt taaagaacaa cattattccg gtcactgagc 37920
ttactaaagc taaaatagaa aatattttta aaagaaggtt gtcgatatata tctctttctt 37980
tccaatccat aatatgaaac cccacatca gatcccatat ttccattta ttagatctaa 38040
tgGCCacaac ctctgctgag taaacattaa tgtaaacatt taattcagta cctttaacat 38100
ttttactttt tactctataa attggcaatg atctgcctct atattcagat cctgattttt 38160
catttacaac ttcttcggta gcaaatggca aaagagttgt ttggccttgaa actgaatcta 38220
tagcatcttc catagaaatc ttgtttaag gctgcccagg catattcaga tacttagttg 38280
agccctttgt tgtgatgata acaatctctt gacctaatct ttttttgaat ttaacttctt 38340
tagctttctc tatttcaaaa tttaacttac ttaaatcaaa agatgtttca acatgattaa 38400
ggtattgctc gcctcttacg agctcaatct tattaaaagc aaagtatatg cctgagatag 38460
tccaaaagtaa taactgaaga gagataaaaa aacttaggta cttatgaatt ttcctaacta 38520
aaaagttcat cttttcataa tcttcattat ttcatcaatt ttacatcaa attcgtcatc 38580
accttttaaa gattcccttaa tgcattcacg caaatgggtt tttagaattt tgccctcaac 38640

agtaataatt gaattctttaa gtgctttttat ttgattaaga atatcaacgc agtattgacc 38700
ttcctcaacc attccttgaa cgcctcgac ctggccctct attcttttta agcttatttaa 38760
ttgctcttta tgacatggat gtttcatttt ataaaaacct cggaactata tacgtaaaaa 38820
ttgtagcga taaaaaatt gctacagatg catagtacag atacctagat ttttttctag 38880
tttctgaaca tgctctgcct agttctgggt cagcagggca aggcataatga taagttttgt 38940
agttgatata tcctgcaatg acaatcatta ctaatgcaaa tagtgttatg taaagcttat 39000
attgagatag ggtaattaag aatggaaata cagttaccag gcttgcaaaa cttgccccag 39060
caccaagagc aacaaaaata gctggcaagg cacaaatat caaagtagaa gaagatgcaa 39120
ataatgaaa aaaattcgac gccttatcat tcatagttat taaatttcta ggatttgaag 39180
atcagacata ttttgccat tagaaagaat ctcttttcg atctcagaga agtgatttc 39240
ttttgcctta gaaaaggcaa tcactacttt tccattatca agatcaacat ctattctttt 39300
gacatcttta tcttttagaa aagttttttga aatacctctt gcacagaaat cacaaccat 39360
tcccttcacg cttacgatcg ctatatcttac atttgagagt ccttcaacaa atgcatcaaa 39420
cctctcaggg ttacatcaa gatctttttcc atcaaccatg gtttcatgta aatgaccttc 39480

atgagaatgc ccatgcatac cactatggtc attctccgcc gcaataaggc cgcttgaaag 39540
caaagtaatt aataataatt tttcatatt tttctccttt taaaatctgt acataaaaatg 39600
taaatcccag ctcgatatctg cgtcatagcc aaactcaaca agggcattac cataaaaaaa 39660
tttcaattct ggatacgtag atgattttatt tgttacagta ttttttttgg ttttaaccat 39720
aaccocatgta tgtaaatcat tgtagtcagc aacgtaagga gcaaatccaa gttgaatata 39780
ttcttcttta tagtctttgc taaatttact gttaatatcc tttaacccaa agcccgcata 39840
ccacttcgg gtttcccagt ctccatggat accataaaaa ttatttttaa tacttccagg 39900
cacaattcct gattgaaaat ataaattcct ttgggatgta aaagtatttt ttctattcac 39960
caataagta aaacgaaaat agttttctgt gtcttataa aatttatttt tagccctttc 40020
taggccaatg gaatatattgt aggttggaga gtagtggaaa taaatagaat catataaggt 40080
atctgattta tacattatgg ttgttccgcc agagtaggat atcgggtcttg catcagcatt 40140
aaagaccatt aaaaaaacta aaatgattat ttatatata cccatggggg gtatcttacc 40200
ataggattat aaattaattt caggcatctc tccaagcgc cattttcaat gaaatcacca 40260
tcagaattcc aaaaataacc atagctatag cagatgcata aaaaatacta actagcccat 40320

atgagagcca gtaaacagct accaatgcta taacagaaaa tactataagc ctgatggccg 40380
tgattataat tggccatttc attgcattgg ctccctgaga tgcaaaaatac agggaaaagcc 40440
catatcccttg aaatacataa catacaccaa ggatttgaat atactgcttt gtaactaata 40500
aggtctctgg atccgatgta aaaatactaa tccaaagatt tgggtgtagt gctaaagcta 40560
gaccgataac agccgaaagc aatccagctg ttgttgcacc aaccatgcct attttttctg 40620
ctcgctctat atttttggca cctatatattg ttccaaaccat ggctgtcata gcagtcccaa 40680
tgccgaacac aattgggatt aataagaact ctacccttga accaattcca taaccggcaa 40740
tagcagacgt tccaaattga ccaatgagtc ctgtcagtaa aagaactgtt gcaactgtca 40800
ttattggcga aagagatgcg ggtaaggcaa cagaaaaaat atcttcgaat aattctcttt 40860
caattgttag tcttttgagc ttaagccta caggcgatga agggctcgaa atttttatta 40920
aagttactaa tgccataaaa cctgaagtaa ctatcatgta ccaagcagat ccacaaagcc 40980
caagtttagg tagaccgaat gatcctaata taaatccagc agagaaaaaa acttgtatatac 41040
ctgcacaaat taccgttagg actgcaggaa attgcatatc tcccatccct cttagggcgg 41100
cggttaagct acctgatagc caaacaacta ttgctcccaa taaatagaca aaacaataag 41160

ctaaagactc ctctaataat gcacctgttc cacctaatat tcttaataga ggctctccaa 41220
aaataaaaaa gataattaaa aaagctaaag ctccaaagca agatatatac aaagaatgcc 41280
ataataattt ttcagcccta ggcttgtctg cggtcccaag acttcttgca atagaagatg 41340
tgaccgctcc tccaagagca ccaaatgcca tctgtgagt cagcatgatt gcaggaaatg 41400
ccaagtaac agcagctaataat ggagttatgc caagttgact aataaaccag aattcagcaa 41460
gaacaaccac agcatttatt aaaaatgcaa cagtattagg tgctgacatt ttaatcagca 41520
aaggaaatat tggatctttt aaaaattgct ctgttctctt gccataaacc ttctcattta 41580
ttaatattta tttgggttaa ggtgaaaga gaacacgttc tcaggattca cgaatcttta 41640
aataataata acagctttta gttaatttaa tttttaaaaa caaccaaatac tattatctgc 41700
tagattttat tttttaaaca tattaatac taatttaatt gcaaaggggt ggccatttcg 41760
gcctgagatc cttggaaact cttgaacct gatccataca ataatggcg aggaattgca 41820
tgaaaatcat taaaatcat ttattttatt caataatctt actatatatt tttagtctag 41880
agacgagctc tcaagctata gaagaagtta ttatcaaagg agactggaga gaaactagtc 41940
tgtcagcaga agactcaagc attgcagtcc ttgatagtaa attaatagaa tcccaggctt 42000

taaagcactt tgaaaacctt tcgtacctag taccaaattt aaattttgct gcaagtgatt 42060
ctagagcaag acattttccag ataagaggaa ttggagaaaag atctggctat gaaagaactc 42120
caaactctgc agtaggttta ctgattgatg atatagactt ttctgggcaa ggtggaattg 42180
ccacaacttt tgatgttgat caaatgaag ttcacagagg cccccaagga gcaagaatag 42240
ggtctagcgc aatggcgggg ctcatatata tttctacaaa agacccaaca gaaagctttg 42300
aagggaaggg tgaaatagtt atggggtcat atggaacttt taataccgga attgctgttg 42360
gcggtcctgt aaatttttaat aaagacctca cttatagggt agctataaag aaagattatt 42420
ctgacggatt taggaagaat atttttctca acaagtctga tacttctaag aaagatgaga 42480
gcacatttag attaaaagt aattgggtaa ctgataacca aaccacttat aagtttctta 42540
tatctcaaat agagttagat gatcctgcag atatatggac tatagatggg agtcttaata 42600
cattatctga taagcctgga atggactctc aaaaaagtaa tgcctatggg gtaaaaaattt 42660
atcatcaatt taaaaaatTT gaatttcaaa gcctatcaag cttaacgaat actgatgtca 42720
ttcttagtta tgatgctgat tggggcaacc ctgagtcgca ttcacccttt atttatgact 42780
atttttcaga aactacaaga aaaagagata cttttagtca agaatttaga cttgtatccc 42840

aatttgcaga taaaaataca gaaaaaagca tcgaatgggt cgttggggct agttttgttag 42900
atataaatga aacaaatgct aaaaaagata ctggtattta tggagatcca tcagacccat 42960
atgggtcctta ttttagtaac tcttcttctt tgagtgactt ttcttcttca agctattctt 43020
tatttggaat tattgattat ttaattaatg aaacaataaa aatttcaatt ggtggaagat 43080
gggagaattt taaatctaatt tatttttgatt cttatgacga atcattttca ccatcaataa 43140
aaatgtcttg tggtaagtgt tcaactagtt aaactcttaa taataattct aatatattt 43200
ttaatatgct taagggtat aaccagggtg gatttaattt aggtcttggt cttgataaaa 43260
attcatcaa tagaaattta tattatgac cagaattttt aactaattac gaagtggaa 43320
ttaatagtaa attttccag tcaaaaattaa atcttgagc agtcctgttt tattctgac 43380
gaaaagatca acaggtctta atttcaacc aggttgatcc ttcagatccc aacacttttt 43440
tatacttaac ccaaaatgct gcagaaggaa ttaataatgg cttagagtta aatatagatt 43500
atgcactgaa taaatctcta ggtatatattg ctaatttttg attgctaaat acagaaaataa 43560
aaaattggat ttcaagacca gatatagaag gtagagaaca ggcgcatgct ccaaaaaata 43620
gtttttcaat aggcatcaat tggaagccaa caaccaatc ttatttatca ttgaatgttg 43680

ttggtaaaag tgagttcttat tactctgatt ccataacaa tacttctgag tcatacaact 43740
taacaaacat taattatgga tatgaacatg gacaatggac ttattcatta tgggcaagaa 43800
atatttttga taagtactat tcagtaagag gtttttactt tggtaacgag gcacctgatt 43860
ttatagatac gctctacaga agacatggag atccaaggca tataggagtc atgggtccaat 43920
atgatttcta actttataag tgaattttgg atggaataag cagctgttgt atttgctatt 43980
atttatattat tacttgctgt aaaacaagac gtaaagtgtt ggtttgcggc tataattagc 44040
tcaatatattat atttcttttat tatgtatgac gcaggcttat atatggaagc ttacctgcaa 44100
atttttttata tcatgatggc cttttatggg cttcaacagt ggagaagtgt cgatactgat 44160
gctcctcaat ttattgttag aacatgggat aaaaggatgc atattaagat cattacatta 44220
atagttatca tgactttaat ctctggattt ttattagaaa aatataccaa tgctatcttg 44280
cctttcatag atggactaac aacatgggga gcaatagttg cgacatatat ggttgccaaa 44340
agacttcttg aaaattggat atattgggtt gttatagatt ttatttctat ctttttattc 44400
atgtccagag gactttttatt aacctcagga ttattcttta ttatcttgt aataatat 44460
tttgggttaca tgtcatgggt aaaaaataaga gacgatatta gtgcagaatc ttcataacaa 44520

tctagatata aaaaatcata atctagagat tattaaaaa attaaatcag gccctgtatc 44580
tgaaatatct atttgtaatt ttgataacat caaagcaatc ttaagagttg atcatccatg 44640
cgcacacaaa ataaatgtag atcgagaaaa tgaaattttt atactcagcc aactaaaaat 44700
ttagatttt agtccagagg ttttattttag tgatttgtct tatggaattt tagtatggag 44760
atataatcgag ggtatttgaat tttcacttgg taaggattcc aatgaagttt ttttaaaaaa 44820
acttggaacc gaattaaaaa aaattcatga tattgatctc ccaaaaagta agaaaaaata 44880
ttagcaat gacataaatt ttacagaaa ttactaaag gaggttcctg aaaatataat 44940
tcttcataga ggatttgatt tatacgacaa actcaataac tctgataaatt atgttctctc 45000
tcataatgac ttgaataaga caaaccttct ttggaggat aggttatattt ttttagactg 45060
ggagtactcg agttttaaca atcctttttt tgataatgcc tcgttatcga atgcttataa 45120
cttatcaaaa gttgatagg caatttttatg gaaagccttat acaaaataatg aatattcagt 45180
atataatgat acaaatctta gagaatggat gcatttttgt cattatttag agtacctgtg 45240
gagtatttcg ctcatacaaa atggaaaaat tgatcagaac accctaaatt taaaaaaatt 45300
agagaaaaa ttaaaaaata ttattaaac aacaagtatt gcgcattctt tagttgttgt 45360

tttactatta tgtaagaaat ttacatttaa atgggggttt gatatgaaaa ttttatgtgt 45420
cttatatgat gatccaaaaa caggtatgcc agaaagggtat gcaagagatg atttaccaaa 45480
gtagataag taccctgatg gaatgacact tccatcccca aaatctatag attttactcc 45540
tggtgagtta cttggttgtg tatctggaga actagggttt cgaaagtttc ttgaagatgc 45600
tggccataca ctagttgtta cttctgataa ggatggagat ggatgtgagg ctgataaaga 45660
attagtagat gctgatatgtg ttatatcaca acctttcttc ccataattatt taacaagaga 45720
caagatgaaa acagcgccta atttaaaaat ggcaattaca gctggcattg ggtctgatca 45780
tggtgatctt caggcagcca tggataatag cgtggatgtt gttgaagtta cttactgtaa 45840
ttctcgttca gtcgccgaac acattgtgat gatgaccta tcgatggttc gtgattatca 45900
taccacaacat agaattgtaa aagagggagg atggaatata gctgatgctg taaaaaggtc 45960
atatgatgta gaggcatgc atgttggtac ttagctgca ggtagaattg gtattgatat 46020
gctaagaaaa atgaaaccct ttgatgtgca tttgcattac tttgataatc ataaactctc 46080
tgatgaaata gaagcagaac taaacctcac ctatcatgat tctgtagagt cattagttgc 46140
tgtgtgtgat gtagttaata ttagttgccc attgcaccc acttattcga 46200

tgatgaaatg attagtaaaa tgaaaagagg tgcataatc atcaatactg ctctgtggcaa 46260
gatttgtgat aaagatgcta ttgcaagagg cttagagtca ggccagctaa gtggttatgc 46320
tggtgatggt tggttccac aaccagctcc aatgatcac gtatggagaa caatgcctaa 46380
ccacggcatg actcctcata cttcaggtag ttcgctatct gctcaaaaca gatatgccgc 46440
tggtggttaga gaaattcttag aatgttattt tgcaggcgaa ccaattagag acccataatt 46500
gattgttcaa aatggtgac ttgcaggtag ggggtgcgac tcgtacacaa aggttacagc 46560
cacagatggc tcagaagagg ccgctaagta taaaaataa gttttagaac ttacttagcc 46620
tttccttaat aatagagtct gcctctgcc taatgctatg cattaattct tccactgtcg 46680
ggatgtcatt aaccagtcca gcaaccattc cacatgacca ggctccaacc tccatagttc 46740
cttcatgcat aatttttgga tagactcctg caacctcatc cacaatatca gcaaaagtta 46800
attcatcacc aagagctttt tctttttcaa tcaatctctc aacagcttca ttattaagaa 46860
ccctttctgt atttgttaat gatctcatga tgagtctagt atctaactca gaagcattta 46920
cgatagcctc ttccacattt tgatgaacag gtgcatcttg agtagcaata aacctagttc 46980
ccatattcat tccctcagca cctagtga ctaggaac caaacttctt ccatctgcca 47040

ttccaccga ggcaacaaat ggtatttcaa gctcatctgc tgctctaggt aagagtatga 47100
aattaggaat gtcattcttct cctgggtgtc cgccacactc aaaaccatct acagaaaaccg 47160
catcgcaacc tattgcttgt gctttttaatg agtgccttac agagggtgcat ttatgaataa 47220
cttttattcc tgcttctttt aaagctggga ggtattctgc aggatttctt ccagcagtct 47280
ctacaacagg aacccttgca tcaataatta cttttatcaa accaggatag tctgggggtg 47340
ttagtgatgg taaaaatgtt aaattaacag caaatggctt attggtcatt tctttgcattc 47400
tggcaatttc atttgctaatt ttctcaggcg taccctgtgt tagaccagta attgtttccaa 47460
gtccaccgc atttgatact gccgtgcaa gctcggcaaa accaacaatgg tgcattccac 47520
cttgaatgat tggatgttct ataccaaaata attcagttat tttagttttc ataatactct 47580
cctattttatc ttccattggg gttaaaaaaat cttcataatg agtttgaaat ctttgcattgc 47640
cacttatcca tctatctctg tcattacctt ttcttttaac atatctaga acctcactat 47700
gaggggttac taaaaacctc tcctcctcta tggcggttaag aacatccttt gcaacaatat 47760
cagcttccat cattccatca actcctgcta caccaggacc atttgcagtc atagcagttc 47820
taactgcttg agggcataaa caggaaacgc cgataacctt atttccatat gtaattttta 47880

tccattcagc aaaactcaca gctgcagcctt ttgtaactgc atatccagcc gcacctagtt 47940
gagttaaaag accggctgct gaggatgtat tcataagata gccttcacct tgtttctatca 48000
tttgaggaag cacatgtttt gcagcatgaa tatgagactg aacatttaca cccatatca 48060
tgtccaatc cgaaagtatct gttcaaaaa atcctggctt tccaccgata cctgcatttg 48120
aacaaaaat atcaatacca ccagaaaaatt cattagcctt ttgtataaca ttataaatgt 48180
cgttttctt ggaaacatca gcacttactg caagcccatt aacactcttg gctgtttcct 48240
ccgccccatt taaattcata tcaacacata ctatagaact tgccccagat gcatagaact 48300
cttcacataa agccttacca attccactgg cagcccctgt aaccacaaact cttttattat 48360
taattttcat aaattgacca cttttttaat tttttattta actttattgc ttagatagtt 48420
tattacaaa caacaatact taatatggca aaatcgcata cccatcttta taaatttatt 48480
gctattatta gtgctagaaa atgaaatact gagatttaat atgaaagaatg ttgttggtat 48540
tggctcgtcc ggggcaatag gaaaagcctt tattgatagc tatatcaaag atgatgatgt 48600
tgaaaaatata ttttcatttt caagaacagg cctttccatt gaggataaaa aactccatag 48660
tttttttatt gatattgagg atgaaactag tatttgtgat gccgcagaga agatagacaa 48720

gtcctcaata gatgaaatta tcgtcgcaag tggaatactt cataataaag attttgggcc 48780
agaaaaaagt attagagatt taaatgcaga taacctttta aaggtcatta aggttaatac 48840
tatcggccca acaattgttg gaaagtattt cattccattg ctaaaataaa aagaaaaaag 48900
cgtcttagca tttttaagt caagagtcgg cagcatttct gataataaaa cagtggttg 48960
gtatgcctat agagcgagta aaactgcact taatcaaatac attaaaagt ttagtattga 49020
attacgaaga accaatccaa atgccattat ttttgggtcta cagccaggaa cagtagatag 49080
tgaattaagc gaacctttta aaagaaatgt aaaagaagg aatttattta ctccagaata 49140
tagtgtattg cagctaaaaa atattattga tacagcaagt ccactctgatt caggcaaaact 49200
aatttcttgg gatggggaag agattcagcc atagtgtgtt atgaatata tttcatatta 49260
gaataaattt ttaggggaaa aaatgagtat taaatattat gactggtcta aatttcaggc 49320
caacactagg ccaaataaag ttgccataag agagctagat aataacaaga tctatactta 49380
cggagaattg gataaaagat catcaaggct tgcatacat ctccaaagt caggaataaa 49440
aaaaggagat cgtattgcga tactatcgct aaattgttca gaattttttg agctagaatt 49500
tgcttgcgga aagattgggg caatagagat accattaaat tggagattaa caaaaccga 49560

gctcagttat attctttaatg atagtgagcc aaaaactcta atttatgaca atcagtttga 49620
agaaatggtg aaagagctaa aagaagaatg taataatttct gaaatcatag ctcttgatca 49680
atttgaccaa gaaagtgatt atgaaaaagt tttgagtaat gcttcaggca ttattatca 49740
ggaagaagtt gatctagaag ataacattat gattatgtat acctctggaa caaccggtca 49800
ccctaagggc gccatgatca cgcacaaaaat gcagctttttt aatgttatta atttaggtat 49860
ttcagcagct gtttcccctg aatcagttcca tttagttgtc ctccctttat ttcatacagg 49920
cggaatgaat tgttattcaa atccaaattct tcatgcaggt ggcgagttaa tattacttaa 49980
agagtttgag cctgggaaag ttctatcaat tatcggcagc tctgactatg gagttactca 50040
tctgtttgca gttccagccc cttatcaatt tatgatgaat catccagatt ttgaaatcaac 50100
aaatttatca ggagttaagt atgctggagt tgggggcgca ccttgtgcag aggctattttt 50160
gaagacttat ataagcaagg gtgttttcgat gcagcaagga tgggggtatga cagaaaactag 50220
tccagggtgct actgggtcttg aatcgtccga ggctgaaaaga aaaataggat ctgctggaaa 50280
accgcttctt cactactgagg tcaagggtggt tggagatgat gggaatgaac tgcctgctgg 50340
agaagtaggc gagatttata ttaaggccc aaatattaca cctggctatt ggaagaaaga 50400

agaggctact agagattctt ttgaagacgg gtggttaaaa acaggtgatg ctgcttactt 50460
tgatgacgag gggttttttat acatagttaga tcgatggaag gatatgtata tctcagggtgg 50520
agaaaaatgtt tatccagctg aagtttsaaaa tgttatctat cagttaccac aaatcgcaga 50580
agttggagtt attggtattg atagccctaa gtggggtgag actggtaaag cctttgttgc 50640
tttaaaagccc gatcatgaat tgacggcaga agaaagtcata gatcattgtt taaaaaatct 50700
agcaaaagtac aaaattccag agaaaagttga gtttattgca gctcttccaa gaaatgctac 50760
aggtaaagtt ttaaaaagaa cattaagaga tatgtaatat ttttaagcca aaaaaaaccc 50820
agcctaagct ggggtttttta ttttaactatt taaagttatc taatgttaag cattagaaga 50880
ttctttaaca gcaacattcc atataattaa accaaaataga atcttgttaa caaagtcagc 50940
aaggttatag ataaggttta agttaagagc tgatccaccg tcacccatca ggtaacctgt 51000
gaaataacct acaggataaa tcgccaacc aaagatgata atatacatca ttgtgttgta 51060
agctgattgc acagcaggac ttgcagtatt acatgcagat tttccttctc cagcccataa 51120
ttcataaatc atgtataccc aagctaaaca cccaataatg aatgcaggcc atgcagccat 51180
gatttcctgct tcacccatgt aaccaaacac aagcataaca agagaaccaa ctagtaattt 51240

cttaaataat gatccagcaa cattagttgc agcagcaaga attaagtaga attcacatat 51300
taatagagga actgttagta accaatcaat gtatctaaat acagttggcg aatcaccagt 51360
ttcaatccat acccctctca tgtacatgta atgccagaaa gcaataccag taacaagacc 51420
agatacagtt aatgatgttt tccattttgc agaaactcta tctctttcaa caaagaaaaa 51480
tacagtagat gctaataaag cagcagtaac taaccaaaaa gaaacaccag tgtaatcact 51540
agcatcaagg tcaccaccac ctgcagcaaa tgtaggaagt gcaataacac tacctaatat 51600
cagtaataat ttcatatata actcctatat taatgacaaa gaagcataat tgcctccccc 51660
ctaattaaat atgaaattta atagtaacca atattaatag aggttacaaa taaaaagcat 51720
ctattttttt aatagaaaaat atgtgtaata aaacaccata ctttaaatta atatttattt 51780
tgcataagta gacttgttta gtatcatatt taacatgaaa gttgcaattt atcctgggttc 51840
cttgacccc atcacaaatg gtcacacgga tattattgat aggggttgcg gactctttga 51900
caaggttgtt gttgcaatag ctaagagtga atcgaaaaac cctcttttta gcctagagga 51960
tagaattaat ttagcccaat ctatttttaa aggaaatgaa aaagtagagg ttgttggttt 52020
tccaagaaag ttaacagttg atcttgcaaa agactatgga gcttgtgcaa ttataagagg 52080

cctacgagca gtttcttgatt ttgaatatga atttcagtta gcaacaatga ataggctcgct 52140
ggctcccaat attgaaagca tttttttaac accaaaagaa agtctcattt atgtatcttc 52200
tagcttaatt aaagaaatat cagacttaaa aggcatata tcgaagttag ttcatcctat 52260
agtagagcag gcacttcgag cgagtagaca ctagctctg acaagcttca caaaaaaaag 52320
tagctctttg attgacgac gtttttaca aa tagtgccttt gcattttattg cagggtctgc 52380
cttctcttcc atatacatatt aattttaaat taaaatatcc tggactacca tcagctgagt 52440
agaagtcttt taatgtagtt cctccaaacct ctattgcttg ctctagtatc ttttttccag 52500
cagctactaa tcttttaca gcactaaat ctaactcatt ggcatttttt agcggatgaa 52560
ttttagctaa gaaaaggctt tcagacgcac aaatatggc gataccgact acattttttt 52620
gattcattaa atagctttta atatttgtt tagagtgact gcaccctgaa aaaaaatcct 52680
ttgcattaaa attttttagaa agaggttcag gcccaagggt ttttattaac ttatgtttat 52740
cgatatcaga agtaagatgc attgatccaa accttctggg atcattataa ataattcttt 52800
cctcatcaaa aattaattca atatgatcat gtttgataaa gaagtttttca ttattttttg 52860
caattctaag actcccagac attcctaaat gaagaataat ttttttatca ttagatagtt 52920

taaaaattat atatttttgct cttcttttcta agcttttcgac aacttgatct tttactgagg 52980
tttcaaaagct atccaccact ttccatctaa ggttttctgtt gtgtatcctt gcttcttttca 53040
gcaaagagcc tttaaaatttt ttaattgctc tcagggttgt ttcaacctctt ggaagtccag 53100
gcatttttaa ctgagtaaat tattaatttt tactatgtca gctggagtaa ttgttcccga 53160
tgcgagccct agtcttagat tggcaagtat gtaatcatat tttgcattag caagattttt 53220
ttccgcgctg tataagtttt tttctgcctg caagagatca acaacgtttc ttgttccaac 53280
tctgtagcca acttgagtcg cttccagggc actagtggc gaaatcactg cttgttttttg 53340
agcatttaca ttgcaacta atgttaaaac atttgaaaac tgggatctga cttcttgaat 53400
aatccttctt tctgtaaata gagtattttc atttgctctt tcatactgtg aatatgcttg 53460
cttccttctt gagttaacgg cgccgccttg aaagagtggc atacttagct gaattgcata 53520
attccttctt cctgttactg atggaactgg aataccttgg ccattgatat taaaaccttc 53580
atagttaaat tggttttgttt cagatttctga ctgacttcca acaatgtcta tcttaggtaa 53640
atgatttgaa gctacacttc ttgcactgct tttcgctgct tttctttctca aatatgctgc 53700
ttttaactgg tagttatttt ccaatgctaa ttcaaacccat gtctcttttg aacttggtgt 53760

tggcagggtca ataagcaaac catctcccaa ttcattttaag ctgaatatatt ctctaccaat 53820
cagagcattt aaagactctc ttgcagaata aagtgacctt tctgtttctaa ttctttgagggc 53880
tttacttaga tcaaatgccca attgagcctc ttgaactcca gttatgggctg ataaccacaac 53940
atcgaaatctt tgttttgctt gatcaagttg ttttttaata gctttttctt cagatatattgc 54000
tgcattttaga ttatcaatag ctctaagtac gccaaaaataa agctcagcag ttcttactaa 54060
aagattttgc tgctcaaatg caaagtctgc ttcagcagca tctgtaagag atttagattg 54120
cctatatatga aaccatgtat ctagtctaaa gagcgggctga gtaaccctgg cagatgtaga 54180
aaaagagtta tattgctgct gcagtttcttt gttttgatag tattcgttcc agttagttga 54240
tccactcaag gtaataactcg ggagaagcgc agctcttctt tgaaccttaa gctctttatc 54300
tgctaaatat gaatatccg ctgctttata tgtaggggtca ttctcaagcg cttcattata 54360
gatatccaaa agactttcag atgaaatatt aaatgagata aaaagtgcta aaaatgattt 54420
tgtgtaaatt ttcataactt attttaacct attaataatg tttagcagtgc aaacattttat 54480
ttttttttaat tttattgaat ttatttttata tttcttagag tagaataatct ctaacaagtt 54540
caataaatatt ttataataaa aaggctttta aaattggcta aaaattcata tgacgctcag 54600

gcaattgaag tcttatctgg actagatcct gtcaaaaaaa gacctgggat gtatacggat 54660
acatctaacc caaatcattt aattcaagaa gttcttgata attcgggtga tgaagctctt 54720
tcaggttatt gttcaaacat aaaagtatct gttctaaaaag atggccttat taaggctctt 54780
gatgatggaa ggggtatgcc aattgatgag cacccggaac ataaagtctt aggtgttgag 54840
ctcatccctt gtaaaacttca tgctggagcg aaattctctg gagatgatta taatttctct 54900
ggtggccctt atggtgttgg agtttctgtt gtaaatgccc tatctgatga attagaggtt 54960
agagtaaaaa gagattctaa agaataccaa atcactttta ataatggaga taagtcttct 55020
gaattaaagc caattgggga agtggggctc agaaattctg ggacatcaat taaattcaaa 55080
ccaaatccta catattttga aactatagag attcagataa aacagcttaa gcatttatta 55140
aaggccaaag cagttctctg tcctggatta acgatagagt tcgttaatga aaaaaagact 55200
gatgataaac aaaagtggta ttttgaagat gggctcaaaa gctatttgat tgatttcttcc 55260
gagggagcag acttggtttt gctagattca attgtatgct ctaaaaaatc tcaagctcaa 55320
gagcttgaat ttgcaatcaa ttggtcatta agaccccaa aaaataaaact cgatgaaacc 55380
tatgtgaatc tcataccaac tgctcagggt ggctcacatt taaatggctt taaggctggg 55440

cttttagatt cattaaaaga attttgtgaa tacagaaatc tattgcctaa aggtttaaaa 55500
attaatgcag atgatgttct taataatgca atttttataa tttcatctaa gcttcagaat 55560
cctcaatttg cagggcaaac caaggaaaaga ctagattcaa aagatcacat gtcgttcgtc 55620
tcaagtacca caaaagacat tttaagtatt tggcttaaca ctcatacaga agagggcgaa 55680
agaatagcag aacttgcaat tatgtctgct cagacgagag caaaagtctc caatatagtt 55740
gaaagaaaga aaacttttag agggccagcc ttacctggaa aactttcgga ctgtaatatg 55800
caggacttaa atgaaacaga gcttttttta gttgaggggg actcagccgg agggtcgcga 55860
aaacaaagcaa gagaaagatc ttccaggca atcatgcctt tgagagggaa gattttaaat 55920
acttgggact tagaaagtgc agaaataata aaatctcagg agataaaaaa cctatcaact 55980
gcaattgggg ttctgccagg aaataatgac ctttcatcac taagatacgg aaaaatttgt 56040
attcttgagc atgctgattc agatgggtctg catattgcaa ctttactttg tgcattgttt 56100
ctaaggcatt ataaaatcttt agttcaagag ggaaggatat atatctcaat gcctcctcta 56160
tatagaatcg attctggtaa agatgttcta tatgcacttg atgataaaca gcgagatgaa 56220
atagttactg aatttaaaaa gaagaagggc aagcctaaag taaacattca aaggtttaaa 56280

ggacttggtg aaatgaatcc acctcaacta agagagactg tgatggacc cc tgctactcgt 56340
cagcttggtc agctttctat cagctcaagc gataatgcaa attctatgat ggacttactt 56400
ttgtccaaaa agaacgcacc agcaagaaaa gaatggcttg aaaagaaaagg gtctctagca 56460
aaaatataaa tatgaaagaa caaataacct caattagcct caagcaatat gctgaagagt 56520
cttatcttaa ctatgcaatg tatgtcatct tagatagagc ttgacctaat attggagatg 56580
gccttaagcc tgttcaaaaga agaatactct atgcaatgtc agagcttggg cttgatgctg 56640
gctcaaaagta caaaaaatca gcaagaactg ttggagatgt tataggaaaa tttcatcccc 56700
atggagacgg cgctgcatac gaagctatgg ttttaatggc tcaaaaatttc tcattcaaat 56760
acccttttgt agatgggtcaa ggtaactggg gttctcagga tgatccaaaa tcttttgctg 56820
caatgaggta tacagaatct aagttaacta aatttgcaaa tcttttaatc tctgaattga 56880
agtctggaac agtcgattgg cagcctaatt ttgatggctc tcttttagag ccagtaattt 56940
ttccagccaa actcccatct atttttattga atggcacttc tggaattgct gtaggaatgg 57000
caacagatat tccatctcat aatattaatg aaattattga tgccacagta catcttattg 57060
ataatccaaa atcacagttg gttgatttac tcaagataat taatggctct gatcttctcaa 57120

ataattcgcc aataattgct agcaaagatg agctgaatga aatttattcg actggaaaag 57180
gctggcttcaa agctcaagcc caatggggcg aggataagaa tcaaatcatt atcaacgcatt 57240
taccttatca agcatctggg tctaaaattt tagagcaaat agctgatcaa atgcttaaga 57300
aaaaaattcc aatgggtggt gatcttactg atgaaggaga ccacaaggag ccagtaaggc 57360
ttgtcataac tttaaaatcg aacagagtaa atgctgaaga tgtaatgaat cacctttttg 57420
catcaactga ttacaaaaa aattatagag taaatatgaa tttgatttca ttgaaagggtg 57480
gaccaaagt tttctcccta gttgatttat tgaagaatg gctagtcttt agaaaaagaaa 57540
ctgtaataag aaaactagaa catagactcg accagggtaa cgatagggtg catatccttg 57600
aggggttatt aattgtttat ttagatttag ataaagtaat aaagattatt agagaatcag 57660
atgagccaaa aaaagatatt attacagctt ttaactttc tgatatccag gcaaatgcaa 57720
tccttgagat taggttaagg caactagcca agctagaaca aatcaagtta gaacaagaaa 57780
gagatacttt ggtagcagaa caaacagata ttgaaaaaat actaagctct aagacaagac 57840
ttaaacaatt aattaaaaat gagcttattg aaataaaaaga tgagtttggg gaagtaagam 57900
aatctccaat aagagaagca acagaagcca aagttttttc tgaagaagaa actcttgtca 57960

ctgagcctat tacagtagtc ttgtctgcag ctggctggat cagaagtgca aaaggccatg 58020
agatagacc cagctcactt tcctatagag gagaagatgt acttcaagat tatggaagag 58080
gaaagagcaa tcaagtttca gtttttcttg actcaaatgg gaaggcttac tcaattgcaa 58140
gtcactctct tccatctgct agaggaatgg gtgatccctat tacagggaagg gtatctgcag 58200
attctggagt aaagtctcatt tcttcattga ttgggaatga tgaagataaa ttcatgatta 58260
tgaatactgc tggatatggc tatatttcag agtttaaaaa tatggtttct aataagaaat 58320
caggaaaagc atttatgaaa atccctcatg aagcagacct tcttaaagcc attaaagtaa 58380
gagacgatca tttgtatata gcagcagttt caaatatgg cagactttta atttttaaga 58440
ttgatgaatt gccaaactctt ggaaagggca aagggaataa aataataaat ataccaaccg 58500
ctaaatttat agcaaaaagaa gagttaatga cccatgcgca acttggttct gaggctagct 58560
ctttaaggat tgaaaagcga aagagattcc tcactttaaa actcaaatg ctagaaaact 58620
atatttctac aagagcaaaa aggggaaata tgcttcaca aggatatagg aaggtagata 58680
aaatgattga agaggttgag ttagaagtta aagaagactg attatagatt tctcaaaaat 58740
ttttaatccc tcatcaacta gctcatttc tattattagg ctggagaaa acctaacagt 58800

agatgcattc gccttaagaa tcattaaccc attatgatgg gatttttttta ttaaatcatc 58860
tatttgaatt ttgctatcct tgcttacttc cactccgacc cataaaaccag cagaagtatt 58920
tttttcaaaa catttgtgct tctcatttat tttattcaac aaatgataa atctaacttc 58980
tttctttaaa acctatttta aaaatgatatt tttagaaatg gtatctataa cttcattacc 59040
tatggcacat gcaattgggc ctctctccaaa agtcgtgccca tgagaccctg ccgacatatg 59100
ttttgaaact ttatctgaag taagaatgcc tcctagtggg aagccattag atataccttt 59160
agcaaaaacaa agaatatcag gagtaatgtt aaattgctca taagcaaaaga gtgttccagt 59220
cctgccaatc ccggattgaa cttcatcaat tatgacaagt gccttatgtt tttttgcaag 59280
tttctttatt ttgctataa attttttatc tgcttttgtt attccagatt gccattgaac 59340
cagtttctaataaaccgctg cagtcttatac tgaaaaaaacc ttttctagat tagtaatgtc 59400
attatatgga tgattttttta tccctctagg cagaggcgca aaaccgtctg taagatgctt 59460
agcttttagct aaagctatac caagcatagt tcttccatga aatgacgttg aaaatgaaat 59520
gacctcattt ttattttttat taactgttga tgagcaaaaat tttcttgcta tcttaactgc 59580
agcctcaatt gattctgccc ctgaattgca aaaaaatact ttatctgcaa aagaattctt 59640

gcataacttt cttgccaaag taactgaggg ctcatattatg tagagatttg ataaatgccca 59700
aagctcctca gattgctttt taagaatttt tattagatct ttatttgaat gaccaaggtt 59760
agtgacggca atgcctgctg tgaatatcaat atatttctta ttatttaggt ccatacatg 59820
cgatccactc gcttttttta caacaaaatc tgcaggagcg taaaaaggca ccataataatt 59880
agtataattct ttaattatct ctttcataaa tatattgtat cgtgggtttca gttcaaaaat 59940
taaatgacca aaaatttcta gtatctttaa aaccaaacag ctactcatt ggctttaata 60000
gaataatatt tatatctagt atatctattg tatgcggtgg catagcattg atatttttct 60060
tttttggagc aacgctcatt ttgccttttg ctggtttaga gcttgggtatt ctattcactg 60120
cattctattt aagtttttaa tggagtgata aaaaagaaaa aatatttatt tctcaagatc 60180
ttgtaactat agaaaaaggc tctaattatg ctgaatatata atgggaagag ttcagggtcat 60240
ttacctcttt ccaggtttca aaagacagaa gagatcttct taagctaagc tttagggtcta 60300
aaggcgagga tgttgagggtt ggcagctttc taaatgaaga tgataaaaaat gtattaatag 60360
aagagctaac tcagatcata gatacatata atcacgattc cttctcaag ccagagcttt 60420
aatctctttt taactcttt tagctttaat tcttttattt cagggatgcc atttctaaaa 60480

tttggaatg attctccctg tattaatggc ataagatatg caattgcttt tggggtaacg 60540
tccattccat ttttagctat aaaagatttt ggtaattttt tttctaaatt agctattttt 60600
gccagtggag ctggttcaat ttccaccta tatttttttag ctttacctct aactattatg 60660
ggcataaac cattcatgcc ttctactgca tactgaactg ccttggtccc aacagccata 60720
gcttgctcta ggtctgtttt ggaggcaata tgctgtgcac ttctttgtag atagtcagaa 60780
acagcccaat gatttttttag ttttaattta tcagtaatta agttggcaat gtaaggcgca 60840
actccaccaa gttgagcatg accaaatgca tctattgttg ctgattcaga aagaaatctg 60900
ttcttattat ttttcaacc ttggatacc acaacaacac agtagccatt ttttttaaca 60960
aacttttaa cttctgctag aaattttttt tgattgaagg ttatttcagg taataaaatt 61020
atatgaggtg catctccttt ttcttttctt gcaagagaag atgcagccgc catccagcca 61080
gcatgcctgc ccataacttc taataaaaa acttttgttg atgttgctgc cattgaccga 61140
acatctaag aggccttctag agttgatgtt gctatatatt tagccgccga cccaatcca 61200
ggacaacagt ctgttaatac caaatcgta tcaacagttt ttgggatggc tatacagtta 61260
attggataat ttaatttttt acttatctgc gaaactttta atgcagtatc agctgaatca 61320

ttaccgccat tgtaaaaaa ataccctata ttgtgcgctt taaaacatc aataagcctt 61380
ttgtactctc tttcgccttga ctcaagatct tttaaacttaa acctacatga cccaaaagcc 61440
ccaccaggcc tatatttcaa tgactctaat gcagatatgg attcttttga tgtatctatt 61500
agttcctcat tcagcgcc caatattcca ttctgcctg cataaatttt accgatgtct 61560
ttatgcttct tggcctctaa aatcagagcg ccagctgtag catttataac tgcggtaacg 61620
ccgccggatt gagcataaaa cgcatTTTTT ttcatTTTTT ctccatcaaa tgtactgaaa 61680
taaaaatatg taaagtatta taactaatat gcgaatacat atcttagga tctgtggrac 61740
ctttatggc gcccttgcta agatacttaa agagtcaggg catgaaatat ctggatcaga 61800
tattcaattt tatcctcca tgcagacta tcttgatagc ttgatattg aaatgatcaa 61860
gggctatgac ataaaaagca tgccagatgc tgacttgat gtgattgga atgcacttct 61920
tagaggaaat gaatccgttg agcatattt gtccaattct ctccgttta agtcaggccc 61980
tgaaatgctt ggagaaaattt taaaaataa aaaggTTTT gctatatctg gaacacatgg 62040
taaaacaacg acttcatata tgttgactca tattatgta gatcagggtta gggatgtcgg 62100
ttttctagtt ggtggaatat caaataatat atctggctct gcattgtctg ggtcagatgg 62160

aacttttgtg attgaggcag atgaatatga ttctgcattt tttgataaaa gatcaaaatt 62220
tattcactat tcaccaagca ctatagtcac caataatatac gaatttgatc atgctgatat 62280
ttttaataat ttagatgaca ttaaaagaca attccatcat ctaattaaaa taattccatc 62340
aaatggaaat gttgtttatt ttgctgatga taaaaatata agagatctta tcgatatggg 62400
aatttgggtca aatcagatag caatcaataa taatgctcat tcaattgagg cagtttattc 62460
tgataagact ctaaaatatg aagaaagtat ttattcatta aatgagttac ctttaatatg 62520
agaacataat tttaaaaact acatttcggc tattttggcg gcaaagacag atggaattca 62580
aattcaggat tctatcaaat cattagctag ctttgatggc gtaaaaaagaa gattagaata 62640
taaaggaaat tttgatggca tagaaaatta tgatgattt gctcatcatc caaccgcaat 62700
agaattttct tctaattgcc tcgtaacaca aaatccatca aaaaaaatc ttgggtctcat 62760
tgagctaggc tcaaacacta tgtctggggg ttctcatggc ttgtctttgg ttgaatctgc 62820
aaaatcttta actcatgtta tctggctaga tcgcaataat gttttgtcag agaattgctag 62880
cattgaatct actaacacta ctgaagattt tatttctgca gcgatatctg ctttctcaga 62940
ttatgatatt gttattttga tgaccaataa agacagccaa aaaatattaa aaccattgt 63000

agatcacttt gaaaaataat aatttaccag tttttccttt aggaatagtc gccctcccag 63060
gtagcatcca atctcttcaa atttttgagc ctaggatat acagatggtt aaaacatgtc 63120
tatccaagaa ccatggattt gtaattgttt ttaatgcaa taatgagtct caaggcgatt 63180
tcacttttc taagaaggga agttttgttg aaattataga ctttaataat ttgccaaatg 63240
gcctttcttg gataactgta aaatctataa ataagtgat aatcagtaat atatgtcaat 63300
tagaagatgg actgcataat gctgatatta aggcacagat agatccagag gtagatgata 63360
aagctgtttt ggcagaatat cctgagatat ctagcattct ttctcagctt gtaaagcatt 63420
ctaagattag tgacctgccc atccaggttg actttggctc tgctgattca gtgcataacc 63480
acttagcagg ccttatacct ctaagctcaa atgagaaaca aaaactatta gaagcattcg 63540
atgcagcaca gcggatgaga attctttcag actatataga aagaatatct actacataaa 63600
ttattttatt taatattatt ggcggcttat tagattttgc tctaaggcaa ttgatggact 63660
taaaaaccgt agggatttct acaaaccctt tttcaatata gtattttgcct ggatagttac 63720
taagcattag cccaataata atagtgaaca aacctggcca ggtaacacca gcatcatgat 63780
tcccccaata ataagggaat aacctaaaat atttttaatc agtaaaaacta aataccacaa 63840

caacgggtta ttggtcttaa atttcgagac atcttttcttt ataaaaatagt cacttggaat 63900
taatccagca agccacctca tactaacgag actaaataaa aaaataaata aagatatatga 63960
gctcagccaa agtataagaa ctttataaga ttcaaaaaaa attatgatgg catttaggggt 64020
atlaaatcc ataattacc acccttttat ttaataatga ttagtatag caatgtctaa 64080
aaataaat tt aaacattaaa taaatatata ataattttgc caaaaaaagc cttaaatgat 64140
atttgactg aaaccctcag tgataataaa gccgaaaaatg ttctatcact tgatatcaaa 64200
ggtattttctt cttttgctga caacatcatt attgcaactg caaatcga taggcattgca 64260
aagtcctttat ccgaaaagt tt agttgaggag ataaaagcta ataaaattag tatcatgggt 64320
gtagaggga agacagaatc aggttgata ttagtagact gtggtgaggt ttagtaaat 64380
attatgaaga atgacataag agagttttat gattagaag gtctttgggg tgaaaaacacc 64440
ctcatcgatt cttcgaaata aatgctatta aatatcataa gtgttgga ctcaccttca 64500
tcttgggagt taacaggaat agaataattac acaaagcaaa tccctaaaga agtaagtcta 64560
aatttttatta acgtaaaagg gcagcaaacac ccaaaaagat ctacagaaga ggtcttaaaa 64620
cttgaatcta agttaatctc gtccaaaata gattctaatg gatataattgt ttgctgggac 64680

tcaagtggcg agtcgttgaa taattttgaa ttttagtaaat tttttgaaaa atccatgctg 64740
gaaaatatga agcttttactt tgtaattgggt ggttcatattg gcataacctcg agatatattctt 64800
gataaatcta ataagataat ctcgatctca agtctcacc tacctcacag gctttttcaag 64860
attgtttcta tagagcaaat ctataggta ttttcaattc tcaaaaatct tccctaccac 64920
aaatgattga tctgaatgag aggtttgttg aaaaaagaag ctttttcaat agattgttgc 64980
ttatatattt tttctttggg gccctgtttt tattcttgct tttcaagacc tattcttttac 65040
aggtttctag ctattctgac tatgaactag cagcactaaa gaataaaaaca aaagaagtgc 65100
tagttcagcc tgttcgagga gtcatctatg acaggaatgg aaatatctta gtttaataatg 65160
tccctacata tgacctaaata atccaacat caaaaattaa aaacttagat gatttcatta 65220
ttaatatctc taaaataata gagctatcag attcagaaat tgaaaacatt attaaaaatt 65280
ttaaaaaggag tgcaacttat aatagagagc taacaattaa aaaagatctc tcgaaagaag 65340
aaatagcaaa atttgaagta agaagctacc aatttcagaa tgcattttata gatgttaggt 65400
atagtagaga aaataaatat ccttacttat tttcacacgc actcggttat gtagggtggtg 65460
taagtaatga taagggtttta tcaattttta aaaatcaaga tttgaagcaa tctgaaaaca 65520

cttttaagta ttcagggtggt tttattgctg gaaaaacagg attggaaaat atttatgatt 65580
cagctttaag agggtcctttt ggaaaaaagt tatttgaagt tgatgcaagg ggaaggccttt 65640
taaaagagct aagtttttgaa aaacctatta atggaaaaag cttttttact catttagatt 65700
taaattcaca aaaaaaagcc ttngaacaaa tgaataatag aagggtgct gttgtgcct 65760
tagagcttaa atctggttct atagttacct acctaaagcac tccaagcttt aatgtaaatg 65820
gtctttctaa tggcatgtcg tcggtagagt tcgaaaaact gattaatgat gtagacaagc 65880
cattttttga tagggccggt caaggtcggt actctccagc ttctacaata aagccagcaa 65940
tagcattgtt tggattataa gaaaaaatag tagactggaa ctttacactt aaagatcctg 66000
gattttttgt attaccagag gatcagagga tttatagggg atggaagaaa ggaggtcagt 66060
gaacaaatga tatgaagggt gcaatcatag aaagtcttaa tacttttttc ttttctcttg 66120
cttataaggc tgatattaat aatttaatca gccatcttct tgagtttggt ttgggcagga 66180
atgtttgtaa agattgtttt aatccagatt ctggattatt gcctacgcct gaatggaaaa 66240
tgaataatct taattttggc tggtttaaag gagataccgt taatttaggt gttgggtcagg 66300
gctatatgag tgccactcca gttcaattag catattactc tgcatttctg gcaaaaaaag 66360

gaaatcttca agagctatca tttgtcgaga gtgacagcct aagcaatact gcttttataa 66420
ataatttaag catagataac tcagactggg atcaaatca ctcaagcatg attggagtta 66480
ttgaagatcc gagaggcact gctaaaaagat taaaaccttt aaaatcatat gttgttgctg 66540
caaaatctgg aacggttgag cttgtcagca cacaaacaaa ggaagattac aaactagtaa 66600
ggcaaaaatat tggtaacaga gaccatgcaa tcattgtggc ttttggtcca atgcctgatac 66660
cagagtatgc agtaagtgtt gttattgaaa atggtgaaag cggcggttct gttgcaggtc 66720
ctgttgctat tgctgtttta aatgagctta tcaataaatg aagaaaaaat tagactttta 66780
aaactttagc atttattttg atcaatattt atttattgag ataaccctgt tgcgggtcat 66840
gggtttattt tttttataca gcgcataca agaagatata agcactgttg ctaagcaagc 66900
tgtattcgtt ggttttggtt tgctgttaata gttttagtt agccaacctg accctgattt 66960
ttataataca ttttctgggt tatttttttg ggggggastt gtattgattt ttctaactat 67020
gatttttggt aaagaaataa atggagccaa aagatggctt gatttaggat tttttaccct 67080
gcaatcctct gagattatta aaatatcatt gccaattttt ttatcatcat atttatataa 67140
taagccactt ccaataagca ctaggcatat ttttattaca ttgatattaa taggttttat 67200

atttgactt gtagctagac aacctgactt gggaaccagt ttagtagtat ttatgtcagg 67260
gggttatgta ttatttttag caggattaag ccgccgttta ttggatctgc aatagcctta 67320
tttttattat ccttacctbt tttatggaat aactttcttag agccatttca acaacagaga 67380
gttttaactt atttgaacc agacgcagac ctttatggta ctgcatggaa tataactcaa 67440
tcaaaaatag caattggatc agcggaatt aatggaaaag gttatcagga gggctcccaa 67500
gcccatcttg atdtttacc agagacagaa acagatttta ttttgctgt tattgctgaa 67560
gagtttggtt ttattggagt ctgtattttg ttatcagtat ttdtctttat atkactcaga 67620
tgtttatatt tagcatttaa tgcaagagat agattttgca ggttaactat aggaggccta 67680
agtttagttt ttgccctctac attatttatt aatttagcaa tggttgttgg tgtagttcct 67740
gtagttggta tgcctcttcc atttatcagc aaaggcggct catctttgct atcctkttat 67800
atagcttttg ggattataat ttctatggca acacataaaa aattaatgca aagatgaaaa 67860
aaattatatt tataacttta atattcacta tttcaattac tgcggattat tcgaatcatg 67920
aagatagcca aatggtaata aatgaactcg tcacaaaaa cggttttgag gaatcttatg 67980
ttactgcaat cttaaaaaat gcaaaaaagc gtgatgagat gcttaaatct gttgctaatac 68040

ctgctgaaa aacaaaaaca tgggatgaat atagggtctat ctttataaaa acaaaaagag 68100
tttcagaagg caaaaaattt ataaaaaaaa atattaatgc tttagagagg gctgagaaag 68160
aatttgaggt tcctaaagaa ataattactg ctattttagg cgtagaaact aattatggca 68220
gcaataaagg tggatacaga gtttagata gtcttactac tttaggtttt gatgacccgc 68280
gaagatctaa tttttttaga agagaactta tagagtctct ccttttaaca agagaaaaata 68340
atttagatat caaaacaaca aaaggctcat atgcgggagc catgggatat gcacaattta 68400
tttcgtcaag ctaccgagcc tatgctgtag attttgatga agatgggttac gttgatttat 68460
ttaattctgt tgatgatgca attgggagta ttgcaaatata tctttatgtt catggatgga 68520
agagagaagg aaagatcgta aaaaaaactt atccaaacaa tgtagaaaa tttataaac 68580
ctcatgagtc tctaacaagg ttcataacctt taatctttaa tgaagatgga aaagatcttt 68640
tttttatgg tgatgataac tttagggcta ttgctaagta taatattagc gatgtctatg 68700
caatggcggt ttattactta tcagaagagt taaaaaatg aaaaaattat tattcacatt 68760
actatcaact tcaatattta ttcaagcaca gagttttgtt ccagattctc ctgagttaga 68820
tctgaagagc tatatcctaa ttgagccaaa caccaatact gttattgcag aatttaattc 68880

agattttggaa atagaaccag ctagcatgac taaaattatg actagctatg ttgtggctga 68940
tcagattgca aatgggtttaa tatctcttga tgacgaagta ctaattagtg aaaaagcatg 69000
gagaatgcaa ggggtctaaga cctttataga agctggaaaa aaagttacgg tatctgatct 69060
tttaaaaggt attatgatc agtcaggcaa tgatgcctct gttgcaattg cagaatatgc 69120
aggcggcact gaaagaggct tcgttgactt aatgaactct tatgctgcct ctttggagat 69180
gaataatacg atcttttcaga attcaacagg gcttccagat gaaaatcatt tttcatcagc 69240
aaaagattta gccaatctga cagctaacta cattaataaa tttccagaag aatatgccttt 69300
atataagcag aagcaattta cttttaacaa tattaagcaa ttaaatagaa ataagctttt 69360
atggagagac gactcagcag atgggtgtaa aactgggtcat acagaagctg ctggatactg 69420
ttagttggc tcagcaaaaa ggggcgggtat gaggcttatt acagttgttg caggaaagtaa 69480
gtcagataat gatcgttttt tatcttccca aagattactt gagtatggat ttagattttt 69540
tacaacacag aagatgtttaa gtgcagaaaa agaatatcag aatattacag tttgggggttg 69600
acaagaaaaag atacttggcg tgggagttct agaggatata tctattactc ttcctagaac 69660
aagtttcaaa aatggttgaag ctgttttataa agttaataac aatatccagg ctcctatcgt 69720

agttggtcaa aaagtcggga ctctagaat aattagtaat gatgagattg ttctagttac 69780
ggatttggtt gctttaaaaa atattgaagc taaagggtttt tttggaagaa tatggtcaaa 69840
gttcgttctt tggatatatta gcttatttgg tctaacagat gaaaacacca cttaaaggcg 69900
ttttaatgg ggcttttgat actgtagata acataaaaaat atcacctttt tctcgagctt 69960
atactttttc tgatagcgta tatgaagtag ttccctttctt taattcaact gcaattgctt 70020
ttgatgatca tataaaaaa ttagaatttt ctgctagtca actagcgatg gccgtagatt 70080
taaaagaagt tataattgaa attaatctt taataacatc atctgaattt agtaatggct 70140
atgtctatta tcaagttact cgaggcggtg atccattaaag gtctcatatt catgagccaa 70200
atttaaaaat agaaaaccttt ggctatgcaa aggcgcactt attccaatgg aaaccattaa 70260
gagtatcagt atgtgatgat attaggtggg ggaggtgtga tattaataca acatcattgc 70320
ttggtaatgt tatgaacatg aatgcagcaa aactagataa ttgtgatgaa gttattatgc 70380
acaaagataa tttgttgacc gaggggtggcg catctaattct tttttttgta aaaaatgatt 70440
caatctgcac ccagctcta aatggaaata tacttcccg cataaccaga gcattactaa 70500
ttaatgaatt aaaaaattat agtatagagg tcatcgaaga taacttcagg ctagaagatt 70560

tatcaaatgc ttcttgcgcg tggctcacia gcgcaacgaa agggctggcc cccattttctg 70620
aaattagtaa cctagaatct cacttggatt tataccatcc tcttttcaaa aagagtgaag 70680
aaatttttaa taagaagtgc cttagttaga agtcagttta ataactatat tatcaagctc 70740
gtaccaaaaa tctttatcac taagaccttt atttgcaaga tcaagctcaa atacattttt 70800
ttggagctgt attaaagtttt gtagtttatg atttttaata aaatgcatat attcagaagt 70860
tttatttttc caaaccctcag atttaattag actatctaca ggattttttac tttgtcttag 70920
attggctgca ctgttaataa tttttaccaac tatccaaact agaagaggtg cataatgatc 70980
ttccgctgat tttattgaat ggataatttt taaagcttgc ttggtattat ttgaaaataat 71040
tttatcttca agctcaaatg gcataaattc tgcagactcc acagaaggtt ttcttttcttg 71100
gccatcacca tctttataag tcagtcttag aagctttact tcatttctgtt gggcaactaa 71160
gtttccagag ttcatctcag atatatcttg cactaaagag ggcctatctt tttcagaaaat 71220
aaaagatagt tgatgcttaa gccaaatctt ttctttcaaaa gatttaagtt ttccacaatc 71280
aatgatgagg gctatctcat ccactctgttt taccactta gttgtgaaat taagctttttc 71340
attgcttgaa ttaataatta ttgcaaatatt atccatatgg ttaatatattt caatttgaga 71400

tattttaatt atttggtcag gtatttttcc ttgatcatga ttaatctcaa ctattatttt 71460
tgagccaaat aaagatccag aagcattttc aataatagtc tgatctattt tatcaaaacc 71520
atcctttatta ataattgttt tttctttgaa ccctttattt gataaatgtt tgagcaaaaag 71580
atcctttagag ttattttttta agacaacctc agatccaaaa ataaagaaga tatttttgtga 71640
ctgatctaaa tattttttgag ttgtcattgc ctacatttc aaaagcatgc acctcaaaaa 71700
ttatttgatc aattaattcc ttctgcattt cagactcaag agatttaatc atttgatctt 71760
gtgcaaatgg attaaattca ttcgatttat atctcttgta gctaaacaagc tttttactta 71820
tttgtttatt tccaatcatt tttatatgta tttctaaaatt aatttctcct tcaagagctc 71880
ttagagaaga gcctccatag atatcatatt tattttaaatt atagtcttgt attagaatct 71940
gattttgacg attctcagct tcttttttctg tattgaagag gattgctaac ttttgctcaa 72000
aagaattagg gacactacta tcaaaccgaa aattaaatag atcttttttta tctagatcta 72060
tatttagttg attaaattgg cagccagtta aagatatcaa caagcaagtg tataaaaaaga 72120
aaggctcttag tttcataacc tcatagtatc tatctgaaga atatatatca aagattaatt 72180
ttaaataaca aaatttatta ttttttcttt aatataaata gtttttttaa tagccgagtt 72240

gtttattgtt gatgccacat tgtcgatagc cagtgctaaa gcttctatat ctttttgctc 72300
aaggtttttg tctatcatta ccttacctct cacctttcca ttacttgaa ctattaaatt 72360
aaattcagaa acttctaata actcctctct aagaacaggc caagaagatt caatttcttc 72420
ttgtgcaaaag tcaaagtaaa aattattcca taaatgttgt gagatatgag gcgcaatagg 72480
attgagagtt ttaagaataa ttattattgc ttcattaaga caatattgat ttgagatcga 72540
cgcattacct tctttaaaag aatcaggat aaagttaact agtccatta tcgaagcaat 72600
ggctgtatta aatgaatc ttgtttcaaa gtcatttgta accttttaa gattattatg 72660
cgattttctt cggagctcta cttcttgtt cgaaggatcc ttgggctcct ctaagtcaat 72720
aaattttcta ttgctgacta agttccatat ttttttcata aatcttgagg caccttctac 72780
agatgactct gaccattcta gactttgctc agggggaggct gtaaacatca tgtaaagcct 72840
tacagtgtca gcgccatact tttctatata agattgagga tctacagtat ttccttttga 72900
tttggacatt ttagcccat ctttaagaac catgccttgt gtaagaagct ttttaaaagg 72960
ctcattgcct tcaactaaat ccatactct aagcgccta tgaaaaaatc ttgaatatag 73020
taagtgcagt atcgcatgct caatacctcc tatatataa tcaacaggca gccaatattt 73080

tgaatttttta tcaaacattt catctgcatt gtcagatgaa gtaaatcttg cgtaatacca 73140
tgatgagtcc ataaatgtat caaaagtatc tgtctctctt ttagtcgat cagaaatatt 73200
ataaaaaatct tcatttttgac ttaagggaat aggtgccgag ttctttttta gctctggaag 73260
cactataggc atatcctttt catcaataac tcttggttca ccattttcat aaacaaccgg 73320
aatagggcaa cccaatatc tttgtctgct cacaccccag tctcttaatc tgaactggat 73380
taactgctca ccaaggtttg cgtctgctaa atctttaata atttcaagag atgcctcatc 73440
agagtccatg ccatcatatt tatcagagtt tattaatttg cttttttgta caactggaag 73500
ttcattatta ttatcaaaac taataacttg aggaatctct agattatatt tagatgcaaa 73560
ttcgaaatct ctttgatcat gagcaggaac acccatcact actccagttc cataatcaag 73620
aagtacaaaa tttccaatcc atacgggtat cttcttctt gtaattggat gaattacatg 73680
cattccacta ttaattccaa gtttttcagc tttagccata tcagcttcag cagctttcac 73740
ttctttgcat ttgttttagaa agtcttttat cgattcatta ttttttgaca accctattga 73800
tatagagtga tttgggtgata ttgctaataa agaaacacca aaaattgtat caggtcgtgt 73860
tgaaaaatcc tttaaagaat catcggaatc ctcaatcatg tatttaattt ctgctccctg 73920

agatttccca atccaattcc ttgtcattgt tttaacattt tcaggccaat ccacctcatc 73980
caaagatgtt aaaagttcct cagcgtagtc tgtaattttt ataaaccatt gatcaatttc 74040
ttttatttca acttgcgccc cagacctcca tcctttttcca tctataactt gctcatattgc 74100
taaaactgtc tcatcaactg gatcccagtt cactaatgat tttttcctat aaaccaaacc 74160
tttatcataa aacttttttaa atattaaactg ctcccactta tagtatcccg gttcacatgt 74220
tctaagctct tttgaccaat catacccaag gcccaaagat ataagttgct gtttcatatg 74280
ttcaatatatt tgattcgtcc aatccttttgg gctaaacttta ttagcaatcg ctgcgttttc 74340
agcaggaagc ccaaatgcat ccagcccat aggctgaaaa acattaaagt tattcattct 74400
tttatactta gatatgacat caccgatcgt atagtttctt acatggcca tatgtaattt 74460
acctgatgga tatggaaca ttgataggca ataaaattt tctctaccat caggattagc 74520
tttaaatgtg tcttctttta accattcatc ttgaatggtt ttctctattt gacttgggtt 74580
atattcagga ttcatttttt taggaactca ttttctaagt aattaatagt atgttttattt 74640
tctttaaagg tatcatcatg cagaatcctt tgatgaagag ggtgatattgt tttaatgccc 74700
tctataaaaa attcatctaa agcactcagc attctcttaa tagccgagct tctagaattt 74760

gcctgagtaa tgatttttgc tagaagtgaa tcatagttag ggggaactct ataccctcca 74820
taaataatgtg agtcatatct tatgccaaaa ccgccagggtg tatgcatttt tgtaattgtt 74880
cctgggatg gttgaaaatt atcaggatct tctgcattaa tctgcattc tagtgagtgg 74940
ccatgaaat taatatcatc ttgatttagc tcaattggca tttcaagagc aattcttagt 75000
tgtgctttta ctaaatcaaa gccagttatc atttctgtta ccggatgttc aacctgaatt 75060
cttgtattca tttctataaa atagaattga ttatcttcgt ataaaaattc aatagtgcc 75120
acccttcgt aatttatttc ttcacataaa ttgacacaag ctattagtgt tttattaaga 75180
gcttcttgat ttataattaag tgctggagct tcctcaatta ttttttgatg tctcctctgc 75240
atactgcaat ccctgtacc taaatgtatc gcctttcctt taccatcacc aactatttga 75300
acttcaatat gtctagggtt cccaataaat ttttctagat aaatggtttc attgccgaat 75360
ccatttttag cttcctgcat tgttatttct gcatgtccaa taagatcttc ctctctttca 75420
acaactctca ttcctcttcc accgccacca gccgttgcct taatcataat tggatagccg 75480
atatctctag caattttttt aaactcatca ttatcagatg gaatctcatc tttataacct 75540
ggaacgattt gaattccaga ttttccagcc agtgtttttg ctgttatttt gtcgccatt 75600

ttttggattg tttctgaggt tggcccaata aatttaaata cacttttttc acacatttct 75660
gcaaaattat gatcttctgc aaggaacca tatccaggat atatagcatt agcacttgta 75720
agttctgcag ctgagagaat ggcagggtata tttaggtagc tttgagttgg agatgcaggc 75780
ccaatacaaa cagtttctatc agagaatctt aaatgtttta gatctttatc agcttcttgag 75840
taaacagata ctgtctctgat accaagctct ttacaagctc ttatagctcg aagagcaatc 75900
tcaccacgat tggcaataag aactttataa ctcattttaa ttgacagtta taattttttg 75960
tccaaatca accggacttc catcttccac gtcaatactt gaaatcttgc catcaaatc 76020
tgatttaatt tcattcatca ttttcatggc ttcaactata cagagcacat cacctacttt 76080
tatattgtct cctactttta caaatggatc tttttcggga cttgggcttc tataaaaagt 76140
tcctactatt ggagagggtaa caacatcccc tatgacagtt tgcttagcct catccaattg 76200
atgagttggt ggtgagatag ctggctcatt tttgacaatg attgggttttg aaacaaatc 76260
ttgattagag ccgttatctc ttgaaattct tactgactca tctccttgac ttacttctat 76320
ttctttttaa tcagactctt gaagcatttc tataagtttt tttattttcc taatatccat 76380
ttatttacc ctgtatttta atatgccctt aagcatcgcc tcttcataac ccttagcacc 76440

aataccagtt atcacttctt cagctatatc agatagatat gagacatgtc taaattcctc 76500
tcttgtatat atgtctgaaa ggtgaacctc atagaatggt atgttgacac caagaaaatgc 76560
atcccttata gctatgctcg tatgagtata tgcagcagga tttattatta caaagtctat 76620
tttttcttca atagtgcctat gtattgattc tattatttca tgctccgcat tactttgaaa 76680
agataattaaa ttgcattgat tttgatttgc taagactaat aaatccttgtt gaatatcttc 76740
aagtgatttg ctgccatata ctcccttttc tctagttccg agaagattga ggttggggcc 76800
attataagc aatatattca tataatagat tttaacagaa atttacagat tttttaggat 76860
tttaaaagata tttttataag agttgatattt ttgcatggg atagcaaaaa ccgccttcag 76920
agcatccctg atagtatat aaaacctccg ataagttcaa tgaagaaattt attcaatgc 76980
ttataataaa ttcattctcta agaattctcag tctcgccaaa aaattcatcc ctatatagat 77040
ctttacttga ctcaagcgtc ttaaatggga gcactttatc tttaaatttg agctctatac 77100
tatccaagta catataataa ccatctcttta ttggccaagt tatatatgct gcttgttttt 77160
caacattggt tgttaaagca aatacttcat tggcttctgg gaccctattt gagtttttcaa 77220
aaagatttgc tgaattcaat tcacctgcaa tcacattgga tagtagtata attattaaac 77280

ttttgatcat gaagacttaa gaataaaccc ctaagtactt aaaataaaga ttttattaaa 77340
taacacggta gttaaaaaatt tatgtttccag acttttcgcac taaaaatttt cttttatatg 77400
cctatttggc ttttaaaaat aattttttat agaaaaagaa cagtaataag agggcatcaa 77460
tttgatgctc aatctgctgc attattgtct ttgcttccaa agaaagattt atctgaatta 77520
ttagatggcg aaattgcaaa agctagaatt actcttgaag aagcaagaat tcaaaaataaa 77580
gtttctttta caccgtctat acaagtcaga aaagtagatc atattttacc aaagcatgat 77640
ttaattctca gagaatacaa gccgcataca gaggatttaa aaaaagttat tctctatttt 77700
catggaggag gttatgtcct taactcagtc aatacacatg atgatatggg ttcatatatg 77760
tcagataaaa taggagtaaa gttttattct ctagactata gactatcgcc tgagagcaaa 77820
tatacctgatt ctcttgatga tgctctagat gctttttctt ggctcatcgg ccagggatat 77880
ggaccaggcg atatttcagt ttgtggagac agcgcaggcg ctcatctagc tgcttctttg 77940
tgccattacc tagctgaaca aaataaaagat ttgcctagta gccagctact aatatatcca 78000
atgtgtgatc cgtcctgttc atctgagtct tataaatttat tatcttcagg atatcttcta 78060
actaaaaaaa ctatgatttg gttttgggat aagcttagaa cttccgaagt aaacaataac 78120

gattcggcct ttaatctctt aaaattttaat tttgaaaaaa ctttaccgcc aacaattatc 78180
gttacagggtg gctttgatcc tctatgcgat gatggagaaa aatatgcata ttattacat 78240
aaaggtaaac ataattgtgaa acaattacat tatccaacaa tgtttcatgg gtttgcata 78300
atgactaat tgaaagcagc gcagatagcc gttgaagatt ttttaaaaga atataagaaa 78360
atactatgag taaaatttta gaagtaagtg acttaagcat caatttttca acaaggggatg 78420
gattgtttaa tgcagtggat aataaagtt ttgatataga aaaaaatcaa accttggcct 78480
tggttggtga gtcaggttct ggtaagtcgg taactgctat gtcaattctt cagctccttc 78540
aaaaaccaca agcatcatat tccaaagagt cttctattaa gtttaatggc gatgagataa 78600
taaatgccaa gtatgaaaag ttactttcct tgagaggaaa tattatatct atgatatctt 78660
aagagccgat gacctcacta aacccttctc acagagtagg taatcagata actgaatcaa 78720
tactacttca ctcaaaaagc tcaaaaaaag atgcaataga tgaagcaaaa aaattaatgg 78780
cacttgttga gattgatgat gttgaaagac ggttctatgc atacctctcat gagctttctg 78840
gagggcagcg acaaagagtt atgattgcta tggcccttgt taataaacct gagctattga 78900
ttgctgacga gccacaacc gctcttgatg taactatcca agcccagata ttagatctca 78960

tgtctaagct taaaaatgaa ctaggcatgt caatactttt tattactcat gatctaggcc 79020
tcgttcaaga attttctgac aatgttttgtg ttatgaagaa tggcaagata gttgaacaag 79080
gaaatactgt tgaggtattt aacaaccctt ctacgaata tacaaaaaaa cttttagatg 79140
cagaacctca gccaaaatta gataaccctg taagtgatga gccataatt gagatcaatg 79200
acttaaatgt ttactactca ataccttcaa ctaatttttt taaaaagaat acttttcatg 79260
ctgttaaaaa tacttccttt agtatttata aaaatactac aattggcctg gtgggggaat 79320
ctggatctgg aaagtcaact ttgggtaagg ctatagcaaa cttagtttct tataagggga 79380
atattaaatt tgaggggaaga gatatcaact caatttccca aaaagaaaaat aaagaattga 79440
aaaaaaatgt ccaaattggtt tttcaagatc cttatgggtc attatcacca cgaatgacag 79500
tggggggat agttggtgaa ggtttagggtg ttcaactttaa gcttacaaaa aaagaaaagag 79560
acgaaaggat agataagggtt ctgtcagatg tcggtatcga aatagtagct aagaataaat 79620
atccgcatga gttttctgga ggccaaaggc agagaattgc aattgctaga tctttaataa 79680
tgaatcctgc ttttatgatt ctgatgagc caacatcagc attagatagg tctattcaaa 79740
ttcaggtaat cgatttatgt aaagagatac aaaatgaata tgggcttact tatcttttta 79800

taagtcatga tttaaagggtt attagatcga tgtcagactt tatttttgtt atgaaaaatg 79860
gagaaatcgt agagtcagga ccttctcaca aggtctttga aagcccagag caagactata 79920
ctaaaaaatt actatcagct gctttaaagt atgcatctga ataattaaat atatymcmtm 79980
tggcaaatag aaartattca aaagagctcg ttgacgggcc taatcaagct gcttctagat 80040
caatgcttag aggagtaggt ttcacatctg aggmtttcac aaaaccattt gttgggattg 80100
cttccacagg agcaaaaagta accccatgca atatgcacat aaatgcactt tcagagatcg 80160
ttgagaaatc agttgatagt tcaggaggaa aggggtgttct ttttaatact attactgttt 80220
ccgatggaat ttctatgggt acacagggtg tgaatatattc tcttgtttct cgagaggtaa 80280
ttgcagattc aatagaaact gttgtgggat gtcttggtta tgatggagt attgctgtcg 80340
gtggttgtga taaaaatatg cctggatgca ttattggaat ggcaagatta aatagaccat 80400
caatatttat atatggtggt tctatcaaac ctagtaaaga aaataccgac tatgttactg 80460
tttgtgagaa aactggagag tactcaaaag gcgatcttaa agaattctgaa ttaattcatg 80520
tagaaaaaat ttccgtaaaa gggcctggat cttgtggggg aatgtatag gcaaatacta 80580
tggcttctgc gattgaagct ctaggcatga gtcttcctgg aagcagcagt caagatgcaa 80640

tttcacacga caaagaagat gattgtttta aggtggcga agcgataatg aatttattag 80700
aaaaagatat taagccttca gatattatga ctaagaatgc ttttgagaat gctataacaa 80760
tggtaatgct tctaggaggt tcaactaatg cagttctgca tttattggcc atggcgcatt 80820
caataggggt tgatttagag ctagatgact ttacaagaat aggaaaaaa acacccgtta 80880
tggcagatct taagcccttt ggttctcatt atatgtctga actcaatgct aatggcggtta 80940
ttcagccact aatgaaaact ttgcttgaga agggattact acacggcaat tgtcttaccg 81000
ttaccgggtca gacgcttgct gaaaatcttt ctggaataaa accttacgag cctgatcaag 81060
agataattaa atcatttgat aatccaatta aatcaaatag ccactcttaga attctgtatg 81120
gcaacttagc gaaagatggt gcagttgcaa aaattacggg taaagaagga acttcctttg 81180
aaggaagtgc tcgtgtattt gattcagaag aagaaggggt taaagcaatc ctatctaaat 81240
ctataaaagc tggagatggt gttgttaatta gatatgaagg gccaaaaagga ggtccaggca 81300
tgagagaaat gctaaaacca acatctgcca taatgggtca aggtcttggc gatcaggtag 81360
cttttataac agacggtcgt ttttcaggag gcactcatgg attgttgttt gggcacatta 81420
ctccagaggc tgcagatggc ggcttaattg caataattaa agatggcgac tctatatata 81480

ttgatgcaga ttctgataag ttaattctta atatttctga ggatgaaatt tcaaatagac 81540
taagtagatg ggtaaaccca aaaacgcctc ccaaaaaagg agtcttagca aaatttgcaa 81600
aaagtgttaa atcagctagt cttggagcgg taacagatta aatatgtatt ttaaaagaaa 81660
atttccaaat agtagactaa gaagaatgcg tctcaattca aacctcagag acttgcttgc 81720
tgaagttagg ttgtctacaa atgatctaat tcagccatta ttataaaaag aaggcttgag 81780
tggaaaagag gctattgaaa gcatgccccaa tattaataga tatggacaag attcaatttt 81840
ttcagaaaata gaagagctac tagagcatga tataaatagc attgccctat ttccagttat 81900
taatgatcc aagaaaaaca gactggaga tgaggctatc aatgcatcta attgatgtg 81960
tgaaactatc agtaacataa agaaacgatt ccctgaaata atctaatat ctgatgttgc 82020
tctagacca tataccgac atgggcatga cggtatttta aaaaatgatt atgttgataa 82080
tgatgagact ttagctgttt taaggaaagca atcactaaca ttagctcagg ctgggacaga 82140
cataattgca ccatcagata tgatggatgg aagaataggc tctataagag aggcattgga 82200
cgaaatggat tataaaaaata caatcctact atcatatgca gcaaaaatata attcaaat 82260
ttatgggtcca ttagagatg ctgttaattc agcttcaagt ctaggcaagt cttcaaaatc 82320

cagttatcaa atgtcaccaa aaaatataaa tgaagctcta catgaagttg ctatggatat 82380
taatgaaggt gcagacatag tcatgggttaa gccaggtag cttatctag atataatttc 82440
aaaagtaaaa gaaaccttca aagtacctac ctttgcatac caggttagt gtgaatatag 82500
catgctaaa ctggcgattg ataaaggatg gcttgaaagc gatgttatgt tagaatcatt 82560
aataaagttt aaaagagcgg gagctgatgc aattctaacg tacgcagcta aagaaatttc 82620
caaggagata actaacaat gagcaatgtg atagaaattc gtgatgaaga aagctttaat 82680
agcgacgtct taaattcaga aaaacctgta ttggttgatt tttgggctga gtgggttgga 82740
ccttgtaaac agcttgcacc aactgttgaa acagttgcag cagaaaaatc agaaacatta 82800
aaggtttgca aaatggatgt tgattcaaat agagagattg ctgctaaata tggaataagg 82860
tcaatccctt cattaattat atttaaaac ggagagcctg caggagttga agtagtgct 82920
ctaaccaaac aacaattaga ggactttata agtacagtag tttaacttg caaagacttc 82980
tttgcatatt gcaagaaata ggattatcat ttcgacttct aaggctgaaa gccactcaaa 83040
acaaaaacc accttttcaa tcataaaaac acaactaga acggaaataa ctaaatgaac 83100
cttactgaaa ttaaaataaa accaataaat gaacttgtag atatagctac tgagcttggg 83160

cttgaggatg ttggaaggct gaaaaagcaa gagataatat ttagaatatt taagcataag 83220
gcttctgaag gtgttgatat ctatggtggc ggagttcttg agattttaaa tgatggtttt 83280
gggtttttgc gatccccaga aggctcttat tgcgctggcg aagatgatat ctatgtttca 83340
ccaagccaaa taagaaagtt tagcctcagg aaggagact cagttgctgg gaagataagg 83400
accctaaaag ataaagagcg ctatttttgca ttaatccaag ttgatactat taatggtgaa 83460
gagccaagaa agactaaaaa caagatttctt tttgaaaatt taactcctct ttttcccaat 83520
gaaagactaa tccttgaaca aggaacaggg tctaatgaag atctttcctc tcgaataatt 83580
gatttgattg ctccaatagg aaaaggtcag cgtggactta tcgtttctcc acccaaggct 83640
ggtaaaacct taatgcttca aagcatagct cattctatta aaagcaataa tccagaaagta 83700
gagcttatag ttctttttgat tgatgaaaga cctgaagagg taacagagat gtcaagaact 83760
gtaaaaggag aggttgtagc tagtactttt gacgagccac ctactcgaca cgttcaagta 83820
gcaaatatgg ttattgaaaa agcaaaaaaga cttgtagaac ataagaaaaga tgtagttatc 83880
ttattagatt ctattactcg tcttggaaga gcataataact cagttcagcc tgcatacagga 83940
aagatatatga gtggtggagt tgactccaat gctcttgaaa ggccaaaaag gttttttggt 84000

gctgctagaa atcttgaaga gggtggaagt ctactattc ttgctactgc tctagttgaa 84060
acaggctcaa agatggatga agttatttat gaggaattca aaggtaacggg taatatggag 84120
attcaccttg aaagaaaaat agccgaaaaa agaatatatc ctgctattaa tattagaaga 84180
tcgggaacaa gaagagaaga ttctactact gctgaagatg aattacaaaag aatgtgggtc 84240
ttaaggaaaa ttttagacga tatggaagat gctcagtcacaa ttcagttcct aatagataga 84300
ttgaaatctc ataaaaacaa cgatgagttc tttacttcaa tgaaaggggg taatggcaag 84360
aagactagat aaagtttttt gccatatcaa tcatcatctt atcagttggg ctttgcgata 84420
cgtgaatttc taattcttta aactcatttt gacatctatc ctttataatt tctgaagcaa 84480
caataaatat tttttttctt aaaactgcat cgtccagatt tgttattaag aatttaagag 84540
tactaaaatt atagattaaa aaaatttcat tattatctgt aacttttggg atttgctcca 84600
gcaaatagat tacttcatag caaactatct catctagact agccttaagt ttttggtgaa 84660
gaaaaccatt tgaatttttcg ccgcaaaaata aaagactctt tcctagaaaaa ttttttctcta 84720
ttaacttaag aattcccttca gatgaatggc tttgtggaaa atgagatttt atgccacttt 84780
caagaagttt attggaagtt gcggggcccaa cagacaaaaa attgtgtgga agatcatcta 84840

aatcgaaaa agattttaaa atatcaagtc cgtacgaagc agcagcttgg ctagtaaaaa 84900
ttaagtttga atatgaatga atatttttaa ttttatcgat agcgattttt gaagggtcga 84960
ttgaattaat tttagaaaaga taaatatattt taagagctat ttcttgagct tcacataaaag 85020
atattaggtt tcctgataaa tttttgggcc tcgtattaat tatcattttt taaaattaaa 85080
gatttcgctc cttccaaaat aaactcatca gcaaataaca taatatctat tgcaatatct 85140
tgaaaggaag atatcttttc tttataaatt ttttcgttgc cttcataaga taatattttc 85200
cctgatattt ttatttggcc atttttatct tcacataaaa tagctattgg agacaggcaa 85260
gtaccttcca tggcagcaac aaaagatcgc tctgcactgg caagaattaa ttctttaggg 85320
tcaccaatat tttctaaaag ctcaataata tcttttttat ctgacagaca ttcaattgct 85380
atataccct gagatgcaga aggtaacatt tcttcaattg agaattcata tgaattttgt 85440
tttaaaccac gccttttttat agcagcctta gcaactacca gcccatcaaa cagaccatca 85500
ttgagtttct ttattctagt agctatgttg cctcttactg gaatagtttt tatatcagg 85560
cttaagtatt taatttgtgc ttttcttctt ggaccagaag taccaatagt tgaatttcta 85620
gcaaattcag aaaatgattt cccatcctta gatagaagca ggtcttcttg agattcccta 85680

ctaaaaacac ttattatttt aaactttgga tcaagctttg ctggaacatc ttttaacta 85740
tgcactgcaa tatctgcttc atcagcctct agagaggact caagtgtaga aatgaatagg 85800
ccctttccac caatttcattg aagaggttta tctgtttggt ctcttcttga tgtcatagga 85860
acaagctcaa ccttaataatt attgatctta gctaataact catctgcaac aaattttgcc 85920
tgatacattg ctagtcttga ttgccttgta gctattctta ttttcatttt gtttccaata 85980
aaagagcctt cacatcgcca aactctttt ctttcaagat agtgaattcc tcaggaattt 86040
ccaaggtagt aaatttacta tactcaaggt agattttaca agatggcttt atttcattct 86100
ttcgaattat agatttttaga acttttaatt catattcttc gccaaaagga ggatcaagaa 86160
gaattaaac aaaacgagat aagtcatttt ttttaatcca actaaatgca tccttaaaaa 86220
aaactttaga ttatcccttg atgccaagaa gttgaatatt ttttgctaac acagaatagt 86280
tttttttatt aagttcaacg aacacaactt tttttgactg tctggatatt gcttcaatac 86340
caagcgcacc cgttccggca aacaaatcaa gacatatata attttcaatc tcaaattgaa 86400
gccaatataa aattatttct ttttaatttat ttgaagttgg tcttaaggaa tccttaaat 86460
cgaaagggtat ttttttacct tttaataaac ctccagtgat tctgatatgt tttttcattt 86520

tttttgcaaa atgggtatgc ctatttatta aataaatata attamaaata catttaaact 86580
aagttataat tcatccatgm gtcctacaag ataatttttag aaaagcaatg aggagctata 86640
tttactctgt cagtgtgatg tcaaatgttg acgagaataa aaagttttagt gctataaacag 86700
tttcttcagt tacctctgtt tctttagatc ctcctagttt gctcgtttgt atcaataaat 86760
cagctggaat tcacaactca ataaaagaag ggtcctctttt ttgcataaac ctttttaata 86820
aaaatcaaga agatatattct aatctatgca gttcatttaa gaccgaaggt gatagattta 86880
atagtgtga ctgggattta agcggcactc cgttttttaa aagtgccag gctaataattt 86940
tttgtactgt tgatcaatta attcatacc acacacatac tattgtgatt ggtcatgtca 87000
caaaactctct tagcgatgag aaaattaata cactgacata tgttgatggt agctatgaat 87060
aaattttcaa aaaatgtatt ttttatttta ataatactca actctttttt tcttgcctct 87120
aacatttttg ctagccagga agagtgtgaa gaaaagccaa gtgttttttat tatctctcct 87180
caagacgggt ttatctcaga atctaataat gtaaaagtct tatttggatc aaaaaatatt 87240
gaaataaatc cagctggcaa aggtgagatt gcaaaaaata aatgttttgc aagcgggcat 87300
caccatcttt tagttaatat cgaagcattg ccagagagct ttatctcttt tgacaagggt 87360

tatttacatt ttggaggagg tcaaaactgaa acaattcttg atcttgatcc tggaacctat 87420
tctctccaac taattcttgg atctttatgtg cataattcaa aaatgcagggt aaataacttt 87480
aaagggtcaag gacccttttt atcagaaaaa ataacaatta cagttaatta gagattagac 87540
cagttataac tttatctaga tgttcatcta agttattagc aattatatatt gcactagaaa 87600
taggaccaga attatcgtca taatcatcac caactacatt cacctttctt attatcccaa 87660
caagaccatg caacatagac caaagagtta tacattttaaa ggcaatcact tcctcaggct 87720
cgtctgctag attagcaaaa cttttttctca tattatcgta tgttccattt gcagatttta 87780
gaagatctgg gtaatcggca aagttaccaa cagctgttc aaacatcaaa tcatatgtat 87840
gtgcattttt taaaccaaac cctatatatt tgcttgcatt agtaacaagt tttttctttg 87900
taattttttt tggattctca aaaaaaacia gctcattaag ttttttaaag ccaacgggtg 87960
caacagcggc atatacacat tcttttgttt caaaatgcct ataaggagca gtttgagaaa 88020
caccactttc ctttgcaagg gatctaatac ttagtttagt gtaaccatct ctatcacaaa 88080
gcctgcatgc gcattctata agttcttctt ttaagtttcc gtgatgatag ttattcataa 88140
tttaatttta atgttgacac tgctaacata tatcatgtta ttatgtatac accgcataca 88200

ttaagttgat tacatatata agtcaaatac agatatatga acataacaaa aataaataac 88260
accgtcatag ctctttttatt aggcaatgcg tttctatcaa atcttgaggc tttagaagtt 88320
cttgaggtta aaatgcttga tgaatatgct gtaactagag aatttcctgg aaagctcatt 88380
ccatcagacc agtctaagct agcatttgaa atacctggaa agataaactc tattaatggt 88440
gatatcggag atgagggtcat cttaggggat gaacttgcct cattggatga tagagaagct 88500
ttagcacaaat taaatcaatc aaaggctaaa tttgatttag ctgaacaagt actagcaaga 88560
tatataaatc tcagagcaga tgggcataatt tctattcaag atcttgataa ggctgaatct 88620
gatcaaatag tagctaagtc gcagtatgat ttttatagag ttaaatttga gcaaaactaag 88680
ttactagctc cttttaatgg agttattcaa aatagatttc tagatacagg atcagtaatc 88740
aatgcaggtg tccaagtttt agaaatttta ggctctagca atgttgaaagc aagaatttct 88800
attccaatga actatatgaa caagggttaag attggagatg agtatgaatt tgatatcaga 88860
ggaatatcta caaagctac gttagagaga ctggctccca tgtcaaccgg aggctccaat 88920
aatagggttag caatttttttag atttgatacc tttttttaatc caggatcaat agtaaagctc 88980
aaactaagca tcaatgagaa agcaaaagggt acatgggttc caattaagtc actgtcccag 89040

tccgaacaag gtatttgggc catatatacc attaacgagc acaagtagt tgtagagat 89100
cttggtgatg ttatttattt tgaagacgaa tatgcttttg tcagcggaac acttaataat 89160
ggtgatttag taatttttagg cggagctcaa aaaattattg aaggaataat aataaataa 89220
aatgaatgtt attaattttc taatagaaa gcctaggata ttatttctaa cttagcatt 89280
tataattactt tctggaattt cttctgggct ttcagttcca attcaagaaa accctgaact 89340
ggctgagaga tggggagggtg ttcgtatttt tcttccccgg gcatccccag aaagaattga 89400
aacagagata gtaaatgatc tagaaatcaa acttagagaa gttgaagaaa tcgatgagct 89460
tgaatcaatt attactcaag gttttttcaac aattgtagtt gaattaaatc aaagtgtacc 89520
tcctatactt attgaagaga cttgggtccaa ggttcaagac aagctcaatc aaatagttat 89580
tcctcaaggt gcagaaatat ttcttgatag aagcagtggc ccgcctatca ctgttcagta 89640
tgctgtaacc tggaaacggca gtggagatgc tccactaata atgatgtcca gactagcaag 89700
ccagctaaaa agaaaaattaa gctcaatagg ctcatctcat caaactgcaa tttttggtga 89760
aacagatgaa gaaatttttaa ttgaactaga ttcatcaaag ctatcttcgc ttggattatc 89820
atttcaagat atcgcaagtg ctattcaagc cctagatgca aaaaaacctt ttggtgtatc 89880

ctcaacaac aattctgagc ttttatatag actcaaagat aatacacaga gcattcaaaa 89940
actctcagaa atacctatca aggttatttaa taaatcagag atcacacagc tagatgatgt 90000
ggcatttatt tcaaaaatcc cggtttctcc tattgaagac atatcttgt ttaatggaaa 90060
ttagttatt tctgttgctg gaaccggatc atttctcaa agagtccatg attatgtaga 90120
acgcgcaaca attgttgtag atgagatgag agaaactctg ccgactgaga tcatataga 90180
ttagtttat gacgaatctg cttacacac taaaaaattt aatgagcttg taaaaagttt 90240
ttcattagca atattttttg ttttagcttt aagtcttttt tttcttggaa ttagatcagc 90300
aataattgtt actcttatcc tgccattttc tatttgctg gttatgattg gttgtagggt 90360
tataggctta cccttgcata tgacatctat cactggaatt attattgcac taggattgct 90420
tatagataat gggattattg ttgttgaaga ttataagaat agaagagcat ctggattaaa 90480
tatcaatgat tcaatttcac aaggactaaa aaacttatgg gctccattag ctgctgcaac 90540
agcaacaacc gtcttctctt ttcttcctat tgttactgga gaagggtcga gtattgaatt 90600
cgtaggcggc atggcaatga cagtcattat gtctataaca tcttcattag ttttggcggt 90660
attaatgggt ccagttctga tgagttatat ggaaaaaatt ccgttcttta aggatgtgga 90720

tattagcaag gaaggggtata gaaatgaaaa aatccttaata aaatataggg ccttttttaaa 90780
ctgggcgttc ttagttccta gaagagcaat catgatatcg ctgcatcgc ctgttcttagg 90840
attctttctt tttaattctt tacctaaga tttctttcct gctcaagata gagatatgtt 90900
tagagttaat atagaactgc cttctaacgc ctcatcactt actacaatgc agagagttaa 90960
ggaaattaga gaagatatc tagatagtga ttttaatttca atagaaaaag attattcgtt 91020
tatacggcaga atgatgccta gagtttttgat gaatgtttgtt ggtggagaag aaaaacaagg 91080
atccaataat attgcgcagt ctgtattttt tgctactgat tattatgaaa tgattgaaaa 91140
ccttcagat ttatcaagaa gactgggttaa aaataaccct gacattatag ttaattattga 91200
tagtttctcg tctggcccc cggttttttc agatgttagt tatgtaattt ttggagatga 91260
tccagattta ctaaaatcac ttggtgagga gctagagcta attattaaca attctcctga 91320
tgtgagtctt acgaaatctg caacttcaaa ctcaataacc aatgttgagt ttgaacttaa 91380
cagctcaaat atttcaactgt ctgggtcaaaa tgccaattat cttgtaaatg aaatgttttac 91440
tgcaaaacaat ggaatatattg ttggcactat gttggattca acaaaagaaa taccagtcag 91500
gctgaaaggg ctgtctaata aaaacaatat tacgggaaat actagtttta taacaatgcc 91560

ctctcaaggt ggtttttgagt attttgatag tttcggaata agctactaa caaacaatac 91620
gtcaacaatt actaggcttg atggcacaag aacaaatgat gttgagggct ggatttggac 91680
aggtacgctc ccgtctgcta ctgaaaaatc tattaaaaa gatgttaaag attttgaatt 91740
aagattgcc aataggctatt catataaaca acttggcgag gctgaaagca ggggcccaatc 91800
tcaagcctca ttatactctt cagctttttat gtattttcatt cttataatag taggcttggt 91860
tatggcgctt aattctttca gagaggctgg tctaattttg tctgttgcac tcttatcaat 91920
tggactatcc tttcttggtt tatttatagg ccagcaaaat tatggattta ttggaactat 91980
aagtgaata gggttaattg gcttatcaat aatgattca attattgtct tatctcacat 92040
aaaaagaag gctgagaaga aatcactaac caaagctgag cttgttgaag ttgttatcag 92100
gtctacacgt cataaatca ctacctctt gacaacactt ggtgggtttc ttccacttat 92160
ttttgcaagt gtattcttca aaccgcttgc ttgggcaatg agcattggag tattaggcgc 92220
gactattaca gccttattat atattccctgc aatgttttatg ataatgagaa aggttaagta 92280
ctagaacaac tttccgagca ataatttctt cctagctttt ttacaataag tgattcatga 92340
gtataggctc cgcaagaatc acatttaac atagagttat caatcttttt ttatgaggt 92400

cctgatatga gagatctgaa aaagaatacc gcagttatta tcaatataaa tatgaccaat 92460
ggaatcaaaa gtaatatcga ttttaataag aacatttttt agtcactaga tgctggctta 92520
ctattttcta taaaattcca agaccagcca gacttatctt cagatttttt tgattgctca 92580
tcttggtagt taatagaaat tattttttta gtatcttcaa gcaattcaat atatcccaaa 92640
gattcatatg aagcctctaa gatcttaaga gctctataat tttcactaga gttaggaata 92700
ttttcaatta cataatttgc tcttcttatt gctgctatat gcgcatcaac actaacgtaa 92760
taatcagctg ccgcaagctc atttcttgca atcatatttc ttaaatagat atttcttgt 92820
ttagcatatg tagaatatg gctatcagga aatcttgta agaattcagt tagttctgaa 92880
aatgattcct tagctcctga gatattctta tttgatagat cagtatctgt cattcttaca 92940
ataaagctat tatctctcgt atagctagaa aggcctttca taaaatatgc ataatacaata 93000
tttggatgcc taggatgaag tcttataaat ttttctgctg cagcatgcgc agcctcagtt 93060
tcagcattca taaattgagc ataaataagc tctacttgcg cctgttcagc atatttgcca 93120
aaaggatatc ttgattcaat tgcttctaata gaatcaatag caccaaaata attttttcca 93180
gccatccttc tttgggcttg atcgtataaa attttttcag gctgttctat ttctgggcca 93240

tcagaattac aactaacc aa taacagagtt acaattggca atactataaa tagttttaaa 93300
atagttttat ctttcattat ttgcacattc tacctgcatt atcgtaattt aagcttgaaa 93360
acatctatct agattaattt ttttataata agttcatatg attgttaaaa atgttccaaa 93420
tgatctatct tcaatgaggc tagacaaagc tacagcagaa atgtttacag attattccag 93480
aactcagata aaaaaatgga tagaagaagg cagagttctt ataaatggag aggtatccca 93540
gccaaagagat aaagtttatg agaattgatca gattgaatta agccctaaag aagaacaaaa 93600
agtatcatgg gaagctcaag atatatgattt tgaaattcac tttgaaaaatg aagattttat 93660
tataattaat aaacctgctg gtttaataat gcattccagg tctggttgct atgatggaac 93720
tctcgcaaat gggctcattt ataaatttcc agaattgatc aatatccaa ggtcgggaat 93780
tgttcatcga ttagacaaag atacttctgg cattctgctc gtagcaagaa atgagtcttt 93840
taggaacttt tttattaatg aaatgcagga gaggagagtt gtaaaaaaat atacgtctat 93900
tgttattggt tctacactag gaagcttttc tatagaagag ccaattggaa gagataaaaa 93960
taatagaacc aaaaatggcaa ttcgagaaga tggcaaatg gctttaacat ttgtaaagct 94020
taaagaaaat attggaaact actctgtgtt ggatataaga atagagacag gaagaactca 94080

tcagattagg gtccatctat catcaaaaaa actaccaata attggagata aaacctatga 94140
cccaagcagg tctattgcaa gagatacccc tgaagagcta attaataatta tccgaggttt 94200
tccaaggcag gcattacatg caacacacct ctcatccaat gaccaaaaaa caaataatat 94260
tttttctttt gatattccca ctccaaatga tatggaggaa ctacttctag aattaaagaaa 94320
attgatctaa tagtaactaa aaacttgttt tttgattaat aaaataatat aaaccttatt 94380
cctaagagtt ttttggttaag aaattgaaat tatctggcg agacatgcta atgcaagcac 94440
ttcatgatga aggtgttgag ctaatctttg gctaccagg tggagccgag cttcatatct 94500
atgatgcaat ttttagacaa gataaaatag atcatatttt agtaaggcat gagcaagggtg 94560
caaccatgc agcagatgga tattcaaggg cgacaggtaa gccaggagtt gtcttagtca 94620
cttctggacc tggtgcaaca aatgctatta caggaatcgc gactgcattt atggattcca 94680
taccaatggt agttatttca gggcagggtg ctagccattt aataggtagt gatgcttttc 94740
aagaaaactga tatgattggt gtttcaagac caattgttaa gcatagctat acagttttta 94800
atgctgaaga aatacctaag ataattaaag aagcttttta tgtcgcaact tcaggcagac 94860
ctggaccctgt tgttatagat atcccaaaag acatgacagc tccggataat ctttttgatt 94920

actcgtatcc tgaagaagcc aagataagat catacaatcc tccgattgag ccagaaaaaa 94980
atcaaataga tagagcagtc gaagctatat tgatatcaaa aaaaccagtt atatatgctg 95040
gtggtggggc aattgctagt aatgccgaaa aagaattact tgaacttaat gaaattattg 95100
atgctcctgt tacaaaatact ttaatgggat tgggtattta ccctgctagt catcatagat 95160
ttcttgggat gttagggatg catggaacat atcaggcaaa tatggcaatg cataatgcag 95220
acttaataat tgctattggc gccagatttg atgacaggat taccaataaa ccatcaaatg 95280
ttgcacctaa tgccaaagtg gttcatctag atgttgatca ctcatctgta tcaaaaaatta 95340
tagaagcaaa tgtagctgtt tttgggcaag taaaaaatc cttaaaatta ataaaagaaa 95400
ctcttgaaaa aaaattagac tcttacgatt ctctcgctct tcagccttg cagatcaga 95460
taaaagaatg gaaatcacta catggtttaa attatgagct ttataaagat gaatctgatg 95520
atcatcccat ttaccccag gctgtagtcc agcatgtcca tgagattaca aatggggaag 95580
catatgtgac ttccgatgtt ggtcagcatc aaatgtttgc tgctcaatat tatcattttg 95640
ataagcctag aagatggatc aattctgggtg gtctaggaac tatgggtttt ggtttgccag 95700
cagcaatggg tgtaaaactc gctttttccaa aagatgaggt tgtttgcatt actggtgagg 95760

gtagtatcca aatgtgcac caagaattgt ctacatgtct tcaatataat ctcccaataa 95820
aaataattaa tatcaataac gaagctcttg gtatgggttaa acaatggcag gatatgaatt 95880
atggaggaag gcactctgaa agtacctatc aaaactcgtt accggatttt ataaaactgg 95940
ctgaatcata tggatcata ggaattaaaa ttacaaaaa ttctgattta agtgaaggct 96000
taaaaaaagc ttttgaaatg aaagataaac ttgtctttgt tgatatattat gtagatcctt 96060
cagagcatgt ttatccaatg caggttgcaa atggcagcct agaaaaatat tggctatcaa 96120
aggatgaaca aacatgatta aaagaaaaact aattttaatt atggaaaaata aaccaggagc 96180
tctagtaaga gtagttggac tgtttcatca aagagggtac aacattgaaa cccttcatgt 96240
agatactgtt aaagactttt ctacttaca atcgatatgt aaaaaaaact tgaaaaccaa 96300
tttgaggata atcaaatatc tagactgacc atagaaacaa tggtttcaga tgaccttatg 96360
aggcagattt tgagacagct caataaatta atagatgtta tagctgtaag caatgaagag 96420
acaacctatt taaaaggagt attattagat gaaaatttat tatgacgaag atgcaaacat 96480
agaaattatt aaagggatga atgtctcaat aattggctat ggatctcaag ggaatgctca 96540
tgcaaaataat cttcatgaat ctggtgtaag tgttactgta ggtttaagag aagggtcttc 96600

ttcatgggca aaggcagaag aagcaggctt aaaagttcaa acagttgctg attcggtaat 96660
ccaagcagat ttggttatga ttttggcacc agatgaattc caaaaaata tatatgaaac 96720
cgaaatcaag ccaaacttaa aaaccagtgc aattcttgca tttgcacatg gctttaatat 96780
tcattttgaa aaaatagttc ctgaagcaac taacagcgtg attatgattg caccaaaaagg 96840
tccaggccat actgttagaa gtactttatac caatggtgga ggcgttccat ctctcatagc 96900
tatatatgaa gatgcttttaa gtgatgaaga ttattcagca aaagatgtag ctctatctta 96960
tgcaaaaagca aatggcggca cgagggctgg tgttcttgaa acatctttta aagaagaaac 97020
agaaaacagat ttatttggcg aacaagcagt ttatgtggc gggcttaccg cttaaatata 97080
agctgggttt gaaactctag ttgaggcagg ctacagtga gagatggcat attttgaatg 97140
ccttcatgaa acaaaaactaa tcacagactt aattcaagaa ggtggcattg ctaatatgca 97200
ttactcaata tcaaatactg ctgagtatgg tgattatgtg agtggaccga aagtaattac 97260
tagcgatacc aaaaagcaa tgaaaggaat actagaaaat atacagtcag gaaaattcgc 97320
agatgacttc ttgaatgact gtcgacaaaag caatgatggc actggcggac ctgttatgaa 97380
aagcaataga gaagctacaa aaattcatcc aatagagtgc gtaggggctg agctaaagtc 97440

taaaatgaag ttcttaaatt cacaaaaatt ggtggataaa gaaattaatt aaaattaatt 97500
aaaaaaaagg tatcttcttc ggaataaactt cgttaagata cgcgtccgca caaagaggcg 97560
gttgttcttt aaaaaatattt ggttactcgt gtgggtgttc aaatacagag aaaaaataat 97620
ttagattttt tataaaaatc aaaaaaacat gatattaatt gaagagtttg atcatggctc 97680
agattgaacg ctggcggtag gcttaacaca tgcaagtcgt gcgagaaaagt atcttcggat 97740
atgagtagag cggcggacgg gtgagtaacg cgtaggaatc tacctagtag aaggggatat 97800
cccgggaaa cccggattaa taccgtatac ctcttcggg agaaaagaagg cctctctttg 97860
aagctttcgc tactagatga gcctgcgtaa gattagcttg ttggtgaggt aaaggctcac 97920
caaggcgacg atctttagct ggtctgagag gacgatcagc cacattggga ctgagacacg 97980
gcccagactc ctacgggagg cagcagtggg gaatatggga caatggggcg aagcctgatac 98040
cagccatacc gcgtgtgtga agaaggcctt cgggttgtaa agcactttaa gcagggagaa 98100
aaagttataa gttaataacct tataaccctg atgttacctg cagaataagc accggctaata 98160
tccgtgccag cagccgcggt aatacggaaag gtgcaagcgt taatcggaat tactggggcgt 98220
aaagcgcgcg tagtggtttt gttaagtgg atgtgaaagc cctgggctca acctaggaac 98280

tgcatccaaa actaactcac tagagtacga tagagggagg tagaattcat agtgtagcgg 98340
tggaatgcgt agatatattatg aagaatacca gtggcgaagg cggcctcctg gatctgtact 98400
gacactgagg tgcgaaagcg tgggtagcga acaggattag ataccctgggt agtccacgcc 98460
gtaaacgatg acaactagct gttgggagac tatgtctttc agtggcgcag ctaacgcctt 98520
aagttgtccg cctggggagt acggccgcaa ggctaaaact caaatgaatt gacgggggacc 98580
cgcacaaagcg gtggagcatg tggtttaatt cgatgcaacg cgaaaaacct tacctactct 98640
tgacatactt ggaggctctt gtaatgagag tgtgcctttt ggaaccaaga tacagggtgct 98700
gcatggctgt cgtcagctcg tgtcgtgaga tgttccgtta agtcggataa cgagcgcaac 98760
ccttaccctt atttgccagc gattcggtcg ggaactataa ggggactgcc ggtgataaac 98820
cggagggaagg tgaggacgac gtcaagtcac catggccctt acgagtaggg ctacacacgt 98880
gctacaatgg gagatacaga cggacgctaa gccgcgaggt ggtgctaata ctaaaaaagtc 98940
tttcgtagtc cggattggag tctgcaactc gactccatga agtcggaatc gctagtaatc 99000
gcggatcagc atgccgcgggt gaatacgttc tcgggtcttg tacacaccgc ccgtcacacc 99060
atggaaatgg attgcaccag aagtagatag tctaacctta gggagggcgt ttaccacggt 99120

gtgcttcacg actggggtga agtcgtaaca aggtagccgt aggggaacct gtggctggat 99180
cacctcctta acgataaaac gcgtttttaaa cgcccacacg agtaaatcaaa tattaataaaa 99240
aagaacattt agatatgtaa aatcatttgtt atgtaatttt ctagtgtata catttatgta 99300
tacataagat cactgcaatt aaaaagtaac atatgcattt atgtgtatgt taataaaagta 99360
attaatatat ttatttaagt tactctcaaa aatgaagata aaacttcaaa aaaaatatgt 99420
aacctttttt aaggttatat gatcaagtaa aggaagagca caaggcggat gccttggcag 99480
cataaggcga tgaaggacgt aataaacctgc gataagcctc ggggagctgg taaataagct 99540
tcgatccgag gattttccgaa tgggaaaaacc caatacacat aagtgtatta tcttatactg 99600
aatacatagg tataagaggc aaacctaggg aactgaaaca tctaagtacc tagaggaaaa 99660
gaaatcaaca gagattccgg tagtagcggc gagcgaacc ggaccagccc ttaagcttat 99720
tttagtccag caaatattc tggaaagttt agccatagta ggtgatagcc ctgtatgcga 99780
aagactaatt taagtgaat cgagtaggtc gggacacgag aaatccttgac tgaacatggg 99840
gggaccatcc tccaaggcta aatactctat gctgaccgat agtgaaccag taccgtgagg 99900
gaaaggcgaa aagaaccccg gcgaggggag tgaaatagaa cctgaaacct tgtgcttaca 99960

agcagtcgga gcagacttgt tctgtgacgg cgtacctttt gtataatggg tcaacgactt 100020
aatctcagta gcaagcttaa ccatttaggg taggcgtagg gaaaccgagt ctaataggg 100080
cgctcagttg ctggaattag accgaaaacc gggtgatcta tccatggcca gtgtgaagg 100140
cgagtaacat cgactggagg cgcgaaccca cttatgttg aaaaatgagg gatgagctgt 100200
ggataggagt gaaaggctaa tcaaaccccg agatagctgg ttctcttcga aaactattta 100260
ggtagtgct cgtgtattac tgtagggggg agagcactgt ttccggctagg gggtcacccc 100320
gacttaccaa accgatgcaa actccgaata cctacaagta tgagcacggg agacagactg 100380
cgggtgctaa cgtccgtagt cgagagggaa acaaccaga ctgtcagcta aggtcccaaa 100440
ttatgattaa gtgggaaaca atgtgggaag gcacaaacag ctaggagggt ggcttagaag 100500
cagccatcct ttaaagaaag cgtaatagct cactagtga gtcggcctgc gtggaagata 100560
taacggggct aaatcataaa ccgaagctac agatcttaaa ttattttaag atggtagaag 100620
agcgttctgt aagcgggttga aggtaaagctg agaggcgaac tggacgtatc agaagtgcga 100680
atgttgacat gagtaacgat caaagagggtg aaaaacctct tcgccgaaaa accaagggtt 100740
cctgtccaac gctaatacgag gcagggtgag gcggcccccta aggcgagggc gaaagccgta 100800

gtcgatggga aacaggttaa tattcctgta ctttttataa ctgcgatggg gtgacggaga 100860
aggttagact agcacggcga cggttgtccg tgttcaaggt tgtaggctgg tgttctaggt 100920
aaatccggaa cgctaaggct gagaactgat aacgaccact ctacgagtgg gaagtagtgc 100980
ataccatgct tccaggaaaa acctctaagc ttcaggttat aagaaaccgt accctaaacc 101040
gacacaggtg gttaggtcga gtagaccaag gtgtttgaga gaactatggt gaaggaaacta 101100
ggcaaaaatag caccgtaact tcgggagaag gtgcgcccgcg tttggtgatg agacttgctc 101160
tctaagctga acgtgggtcga agataccagg tggctgcgac tgtttactaa aaacatagca 101220
ctctgcaaac tcgtaagagg aagtatatggg tgtgacgcct gcccggtgcc ggaagggttaa 101280
ttgatgggt tagcttatgc gaagctcttg atcgaagccc cggtaaacgg cggccgtaac 101340
tataacggtc ctaaggtagc gaaattcctt gtcgggtaag ttccgacctg cacgaatggc 101400
gtaacgatgg ccacactgtc tccaccatag actcagtga attgaaatcg ctgttaagat 101460
gcagtgtacc cgcagctaga cggaaagacc ccgtgcacct ttactatagg ttcgccactgg 101520
actttgacct tacttgtgta ggatagggtg gagactttga agcagagacg ccagtctttg 101580
tggagtcac cttgaaatac cacccttgta agattgaagt tctaacctag gtccattatc 101640

tggatcaggg acagtgcgtg ctgggtagtt tgactggggc ggtctcctcc taaagagtaa 101700
cggaggagta cgaaggatc cttatcacgg tcggacatcg tgaggtaatg ataaaggcag 101760
aaggatgctt gactgcgaga tcgacggatc gacaggtag gaaactaggt cttagtgatc 101820
cggtggttct gaatggaagg gccatcgctc aacggataaa aggtacgccg gggataacag 101880
gctgataccg cccaagagtt catatcgacg gcgggtgtttg gcacctcgat gtcggctcat 101940
cacatcctgg ggctggagca ggtcccaagg gtatgggctgt tcgccattta aagtggtagc 102000
cgagctgggt ttagaacgtc gtgagacagt tcgggtcccta tctgctgtgg gcgtttggag 102060
atttgaggga agctgattct agtacgagag gaccgaattg gacgaacctc tgggtgtccg 102120
gttgtcacgc cagtggcatt gccgggtagc tatgttcgga aaggataacc gctgaaagca 102180
tataagcggg aagcctctcc caagattaaa tctcccagag actttatgtc tcctaaagag 102240
tcgtcataga ctatgacgtt gataggcaag atgtgtaagc gctgcgaggc gttgagctaa 102300
cttggtactaa taactcgtga ggcttgatca tgtaacctta agcaaggttc ataatttgag 102360
taaaacattg tagtgagaat taaaaataa aaagttacat accagtttgc ctgatgacaa 102420
tagcaacttg gaaccacctg atcccatctc gaactcagaa gtgaaacgag ttaacgccaa 102480

tggtagtga ggggtctccct gtgtgagagt aggaaatcgt caggcttttt tctttaaggc 102540
ttccagtta ctggaagcct ttttttttat ctcaagtata ataccagat gattattgga 102600
ttaacaggag gcattggttc tggcaaatct gccgctgcag acttctttat tgatttaggt 102660
atatacgtct tagatgcaga tcaagttgct aaagaagctt tatctacaaa ttctcctgga 102720
tatactgatt ttatttctca atttggtgaa gtgtatttaa ataataatcg tgaggttgat 102780
aggctgaaat tgcgcgaact tatttttttca aatccttcaa aaaaaaaggga tcttgagaat 102840
attattcatc ccatagttcg gtctgctatt agtaatttta ttattacatc aacatcacca 102900
tattctattg ttatggtgcc actcattttt gaaacaaatt catataaaaa ttacgataag 102960
attattactg ttgactgtga tttagaactt cagatagtaa gggcctcaag tagagatgct 103020
caaaataaat cgcagattaa gaatattatt aataagcaag cctctagaga ggagaggcta 103080
agtattttctg atgacgtact tatcaataac agcacccctat ctgatctaaa aaaacaagtt 103140
aatgtttttac atactaaata tatggagtta ttaaatgagt agttgcccta gatgtgaaaa 103200
acctgtcaaa ctttctactg acaatattta tagacctttc tgctctgaga aatgcaaaact 103260
tatagatttt ggtgattggg ctaatgaaga taataaaatt tcaagaccaa ttcaatctga 103320

agattttttac gaagattaaa tttaagaaag tctccattca ccactttcaa tcatcgggtt 103380
ggcttttttta tacttcattt cttgcgtatc ttgaccggtta gtaatttttta caagatcatt 103440
tctgcctaac ttaggttcgt ttcttgtcac agtacttggt tttactatag gagtggcctg 103500
tttttcattt tgaagatat cagaattaat ttccctctttt tccaatttta attccttgaga 103560
actattttgt ttgtttatgc tctctagctc agattctgtc gataattgta atgagaaaaag 103620
tattcttata gtttcaacat ctatttccga aagcatagat tcaaacattg aatatgcctc 103680
tctttttgaat tcatttttttg gattttttttg agcataagcc cttagacca cactatttct 103740
taaatgatct atttctgata aatgctcttt ccaatggaca tcaagtactt gcagcataac 103800
ctgctttctca agaagcaatc tattttcacc aagatcacta aactttttttg aatattttatt 103860
ttttgcttgc agaacaattt cttcgggcaat tgtatttggt acaagttttt tattactatt 103920
aattttattg gctatatctg tttctagtcc atagctctct tttaaataat catcgagctg 103980
cctacttttc cattgagact ctattgattc ttcagggtaca tataaattag atattccttt 104040
gaattgctgc tcgatgagt actcaatggt actgctaata tcttcctctt ctagtaattg 104100
attccttaaa gaatatatag cttgtctcttg atcatttgat acatcgtcgt actctaaaaag 104160

atctttcctt gcatcaaatg ttctgtctctc aattcttttt tgcgcatttt ctattcctct 104220
tgaagcatt ttagcttcaa tatgatcatc tcccatgcc agcctttcaa aagtagccct 104280
tcttccatcc gaaataaaaa gtctcaagag atcgtcttct aaagataaga agaattttga 104340
ataaacctgga tcaccttgtc tgcctgatct cccctttaac tgattatcta ttcttcttga 104400
ttcatgcctt tctgtttccaa gtatatgaag cccaccagat tctataaactt ttttattatt 104460
ttctttccac tctatatcgc ttgtgtcttc tttttttccg ccaagaacta tatccgtccc 104520
ccttcctgcc atattcgtag caatagttac cattccaggt ttacctgcat tggcaattat 104580
ctctgcttct ttttcgtgat gcttggcatt taaaatttga tgaggatat tttgtttatt 104640
taaatatgct gatacttctt ctgaagattc aaccgaaact gttccaacaa gaatgggaga 104700
agatcttttt cttaattggt caatttcttc aattagagct ttatatattcg attctgttgt 104760
aagaaaact aagtcattaa gatcagctct agccatagga acattttgttg ggatgatgat 104820
gacatttagg ccatagat tt gactaaactc tactgttca gtatctgctg ttcctgtcat 104880
cccagaaagt tttttaata atctaaaaaa gttttggaat gtggtggatg ctagtgttg 104940
agactctctt tggatagcaa cattttctt gcattccagt gcctgggtgaa caccttcaact 105000

cattcttctt ccgggcattg ttctacctgt atgctcatca atcaaaagaa cctcaccgtt 105060

cctaaccaaa taatccacat tcttttttaa taagaagctt gctctaagtg ttgcttgaac 105120

aaatttcata atttttaaat tagaaacaga gtaagcccat ctgaggctcc aagccgattc 105180

cagc 105184

CLAIMS

What is claimed is:

1. A proteorhodopsin gene, comprising an isolated DNA sequence for encoding a proteorhodopsin protein.
2. The proteorhodopsin gene of claim 1, wherein said proteorhodopsin gene is retrieved from a genomic fragment of a sample of naturally occurring bacteria.
3. The proteorhodopsin gene of claim 2, wherein said naturally occurring bacteria are marine proteobacteria.
4. The proteorhodopsin gene of claim 2, wherein said naturally occurring bacteria are SAR86 bacteria.
5. The proteorhodopsin gene of claim 2, wherein said naturally occurring bacterial genomic fragment is retrieved from a recombinant DNA library.
6. The proteorhodopsin gene of claim 5, wherein said naturally occurring bacterial genomic fragment is retrieved from a bacterial artificial chromosome library.

7. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone BAC31A8, said proteorhodopsin gene is Sequence ID No:4 and said proteorhodopsin protein is Sequence ID No:5.
8. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone BAC40E8, said proteorhodopsin gene is Sequence ID No:8 and said proteorhodopsin protein is Sequence ID No:9.
9. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone BAC41B4, said proteorhodopsin gene is Sequence ID No:10 and said proteorhodopsin protein is Sequence ID No:11.
10. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone BAC64A5, said proteorhodopsin gene is Sequence ID No:12 and said proteorhodopsin protein is Sequence ID No:13.
11. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone HOT0m1, said proteorhodopsin gene is Sequence ID No:14 and said proteorhodopsin protein is Sequence ID No:15.

12. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone HOT75m1, said proteorhodopsin gene is Sequence ID No:16 and said proteorhodopsin protein is Sequence ID No:17.
13. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone HOT75m3, said proteorhodopsin gene is Sequence ID No:18 and said proteorhodopsin protein is Sequence ID No:19.
14. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone HOT75m4, said proteorhodopsin gene is Sequence ID No:20 and said proteorhodopsin protein is Sequence ID No:21.
15. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone HOT75m8, said proteorhodopsin gene is Sequence ID No:22 and said proteorhodopsin protein is Sequence ID No:23.
16. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone MB0m1, said proteorhodopsin gene is Sequence ID No:24 and said proteorhodopsin protein is Sequence ID No:25.

17. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone MB0m2, said proteorhodopsin gene is Sequence ID No:26 and said proteorhodopsin protein is Sequence ID No:27.
18. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone MB20m2, said proteorhodopsin gene is Sequence ID No:28 and said proteorhodopsin protein is Sequence ID No:29.
19. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone MB20m5, said proteorhodopsin gene is Sequence ID No:30 and said proteorhodopsin protein is Sequence ID No:31.
20. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone MB20m12, said proteorhodopsin gene is Sequence ID No:32 and said proteorhodopsin protein is Sequence ID No:33.
21. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone MB40m1, said proteorhodopsin gene is Sequence ID No:34 and said proteorhodopsin protein is Sequence ID No:35.

22. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone MB40m5, said proteorhodopsin gene is Sequence ID No:36 and said proteorhodopsin protein is Sequence ID No:37.
23. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone MB40m12, said proteorhodopsin gene is Sequence ID No:38 and said proteorhodopsin protein is Sequence ID No:39.
24. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone MB100m5, said proteorhodopsin gene is Sequence ID No:40 and said proteorhodopsin protein is Sequence ID No:41.
25. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone MB100m7, said proteorhodopsin gene is Sequence ID No:42 and said proteorhodopsin protein is Sequence ID No:43.
26. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone MB100m9, said proteorhodopsin gene is Sequence ID No:44 and said proteorhodopsin protein is Sequence ID No:45.

27. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone MB100m10, said proteorhodopsin gene is Sequence ID No:46 and said proteorhodopsin protein is Sequence ID No:47.
28. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone PALB1, said proteorhodopsin gene is Sequence ID No:48 and said proteorhodopsin protein is Sequence ID No:49.
29. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone PALB2, said proteorhodopsin gene is Sequence ID No:50 and said proteorhodopsin protein is Sequence ID No:51.
30. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone PALB5, said proteorhodopsin gene is Sequence ID No:52 and said proteorhodopsin protein is Sequence ID No:53.
31. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone PALB7, said proteorhodopsin gene is Sequence ID No:54 and said proteorhodopsin protein is Sequence ID No:55.

32. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone PALB6, said proteorhodopsin gene is Sequence ID No:56 and said proteorhodopsin protein is Sequence ID No:57.
33. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone PALB8, said proteorhodopsin gene is Sequence ID No:58 and said proteorhodopsin protein is Sequence ID No:59.
34. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone PALE1, said proteorhodopsin gene is Sequence ID No:60 and said proteorhodopsin protein is Sequence ID No:61.
35. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone PALE6, said proteorhodopsin gene is Sequence ID No:62 and said proteorhodopsin protein is Sequence ID No:63.
36. The proteorhodopsin gene of claim 2, wherein said genomic fragment is retrieved from a clone PALE7, said proteorhodopsin gene is Sequence ID No:64 and said proteorhodopsin protein is Sequence ID No:65.

37. The proteorhodopsin gene of claim 1, wherein said proteorhodopsin gene is amplified from a genomic fragment by polymerase chain reaction.
38. The proteorhodopsin gene of claim 37, wherein said polymerase chain reaction is performed by primers with Sequence ID No:2 and Sequence ID No:3.
39. The proteorhodopsin gene of claim 1, wherein said proteorhodopsin gene is derived from a marine environment and placed in an expression vector for producing said proteorhodopsin protein in a host.
40. The proteorhodopsin gene of claim 39, wherein said host is an artificial membrane system.
41. The proteorhodopsin gene of claim 39, wherein said host is a bacterium.
42. The proteorhodopsin gene of claim 41, wherein said host is a cell membrane preparation of said bacterium.
43. The proteorhodopsin gene of claim 39, wherein said host is an eukaryote.
44. The proteorhodopsin gene of claim 43, wherein said host is a cell membrane preparation of said eukaryote.

45. A method of retrieving a proteorhodopsin gene, comprising the steps of:
- (a) providing a sample of naturally occurring bacteria;
 - (b) extracting a genomic fragment of said sample of naturally occurring bacteria; and
 - (c) amplifying said proteorhodopsin gene from said genomic fragment using polymerase chain reaction.
46. The method of claim 45, further comprising the step of creating an expression vector containing said proteorhodopsin gene.
47. The method of claim 45, wherein said naturally occurring bacteria are marine proteobacteria.
48. The method of claim 45, wherein said naturally occurring bacteria are SAR86 bacteria.
49. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is retrieved from a recombinant DNA library.
50. The method of claim 49, said naturally occurring bacterial genomic fragment is retrieved from a bacterial artificial chromosome library.
51. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone BAC31A8, and wherein said amplified

- proteorhodopsin gene from said clone BAC31A8 is Sequence ID No:4 and encodes a proteorhodopsin protein according to Sequence ID No:5.
52. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone BAC40E8, and wherein said amplified proteorhodopsin gene from said clone BAC40E8 is Sequence ID No:8 and encodes a proteorhodopsin protein according to Sequence ID No:9.
53. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone BAC41B4, and wherein said amplified proteorhodopsin gene from said clone BAC41B4 is Sequence ID No:10 and encodes a proteorhodopsin protein according to Sequence ID No:11.
54. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone BAC64A5, and wherein said amplified proteorhodopsin gene from said clone BAC64A5 is Sequence ID No:12 and encodes a proteorhodopsin protein according to Sequence ID No:13.
55. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone HOT0m1, and wherein said amplified proteorhodopsin gene from said clone HOT0m1 is Sequence ID No:14 and encodes a proteorhodopsin protein according to Sequence ID No:15.
56. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone HOT75m1, and wherein said amplified

proteorhodopsin gene from said clone HOT75m1 is Sequence ID No:16 and encodes a proteorhodopsin protein according to Sequence ID No:17.

57. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone HOT75m3, and wherein said amplified proteorhodopsin gene from said clone HOT75m3 is Sequence ID No:18 and encodes a proteorhodopsin protein according to Sequence ID No:19.
58. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone HOT75m4, and wherein said amplified proteorhodopsin gene from said clone HOT75m4 is Sequence ID No:20 and encodes a proteorhodopsin protein according to Sequence ID No:21.
59. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone HOT75m8, and wherein said amplified proteorhodopsin gene from said clone HOT75m8 is Sequence ID No:22 and encodes a proteorhodopsin protein according to Sequence ID No:23.
60. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone MB0m1, and wherein said amplified proteorhodopsin gene from said clone MB0m1 is Sequence ID No:24 and encodes a proteorhodopsin protein according to Sequence ID No:25.
61. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone MB0m2, and wherein said amplified

proteorhodopsin gene from said clone MB0m2 is Sequence ID No:26 and encodes a proteorhodopsin protein according to Sequence ID No:27.

62. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone MB20m2, and wherein said amplified proteorhodopsin gene from said clone MB20m2 is Sequence ID No:28 and encodes a proteorhodopsin protein according to Sequence ID No:29.
63. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone MB20m5, and wherein said amplified proteorhodopsin gene from said clone MB20m5 is Sequence ID No:30 and encodes a proteorhodopsin protein according to Sequence ID No:31.
64. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone MB20m12, and wherein said amplified proteorhodopsin gene from said clone MB20m12 is Sequence ID No:32 and encodes a proteorhodopsin protein according to Sequence ID No:33.
65. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone MB40m1, and wherein said amplified proteorhodopsin gene from said clone MB40m1 is Sequence ID No:34 and encodes a proteorhodopsin protein according to Sequence ID No:35.
66. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone MB40m5, and wherein said amplified

proteorhodopsin gene from said clone MB40m5 is Sequence ID No:36 and encodes a proteorhodopsin protein according to Sequence ID No:37.

67. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone MB40m12, and wherein said amplified proteorhodopsin gene from said clone MB40m12 is Sequence ID No:38 and encodes a proteorhodopsin protein according to Sequence ID No:39.

68. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone MB100m5, and wherein said amplified proteorhodopsin gene from said clone MB100m5 is Sequence ID No:40 and encodes a proteorhodopsin protein according to Sequence ID No:41.

69. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone MB100m7, and wherein said amplified proteorhodopsin gene from said clone MB100m7 is Sequence ID No:42 and encodes a proteorhodopsin protein according to Sequence ID No:43.

70. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone MB100m9, and wherein said amplified proteorhodopsin gene from said clone MB100m9 is Sequence ID No:44 and encodes a proteorhodopsin protein according to Sequence ID No:45.

71. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone MB100m10, and wherein said amplified

proteorhodopsin gene from said clone MB100m10 is Sequence ID No:46 and encodes a proteorhodopsin protein according to Sequence ID No:47.

72. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone PALB1, and wherein said amplified proteorhodopsin gene from said clone PALB1 is Sequence ID No:48 and encodes a proteorhodopsin protein according to Sequence ID No:49.
73. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone PALB2, and wherein said amplified proteorhodopsin gene from said clone PALB2 is Sequence ID No:50 and encodes a proteorhodopsin protein according to Sequence ID No:51.
74. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone PALB5, and wherein said amplified proteorhodopsin gene from said clone PALB5 is Sequence ID No:52 and encodes a proteorhodopsin protein according to Sequence ID No:53.
75. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone PALB7, and wherein said amplified proteorhodopsin gene from said clone PALB7 is Sequence ID No:54 and encodes a proteorhodopsin protein according to Sequence ID No:55.
76. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone PALB6, and wherein said amplified

proteorhodopsin gene from said clone PALB6 is Sequence ID No:56 and encodes a proteorhodopsin protein according to Sequence ID No:57.

77. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone PALB8, and wherein said amplified proteorhodopsin gene from said clone PALB8 is Sequence ID No:58 and encodes a proteorhodopsin protein according to Sequence ID No:59.
78. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone PALE1, and wherein said amplified proteorhodopsin gene from said clone PALE1 is Sequence ID No:60 and encodes a proteorhodopsin protein according to Sequence ID No:61.
79. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone PALE6, and wherein said amplified proteorhodopsin gene from said clone PALE6 is Sequence ID No:62 and encodes a proteorhodopsin protein according to Sequence ID No:63.
80. The method of claim 45, wherein said naturally occurring bacterial genomic fragment is in a clone PALE7, and wherein said amplified proteorhodopsin gene from said clone PALE7 is Sequence ID No:64 and encodes a proteorhodopsin protein according to Sequence ID No:65.
81. The method of claim 45, wherein said polymerase chain reaction is performed by primers with Sequence ID No:2 and Sequence ID No:3.

82. The method of claim 45, further comprising the step of providing a host.
83. The method of claim 82, wherein said host is an artificial membrane system.
84. The method of claim 82, wherein said host is a bacterium.
85. The method of claim 84, wherein said host is a cell membrane preparation of said bacterium.
86. The method of claim 82, wherein said host is an eukaryote.
87. The method of claim 86, wherein said host is a cell membrane preparation of said eukaryote.
88. A light-driven energy generator, comprising:
- (a) a proteorhodopsin protein;
 - (b) a host to correctly fold said proteorhodopsin protein in said host, thereby creating an integrated proteorhodopsin protein; and
 - (c) a source of retinal to bind covalently to said integrated proteorhodopsin protein, thereby creating a light absorbing pigment.
89. The light-driven energy generator of claim 88, wherein said proteorhodopsin protein is encoded by a proteorhodopsin gene retrieved from a genomic fragment of a sample of naturally occurring bacteria.

90. The light-driven energy generator of claim 89, wherein said naturally occurring bacteria are marine proteobacteria.
91. The light-driven energy generator of claim 89, wherein said naturally occurring bacteria are SAR86 bacteria.
92. The light-driven energy generator of claim 89, wherein said naturally occurring bacterial genomic fragment is retrieved from a recombinant DNA library.
93. The light-driven energy generator of claim 92, wherein said naturally occurring bacterial genomic fragment is retrieved from a bacterial artificial chromosome library.
94. The light-driven energy generator of claim 89, wherein said genomic fragment is retrieved from a clone, wherein said clone is a member of the group consisting of BAC31A8, BAC40E8, BAC41B4, BAC64A5, HOT0m1, HOT75m1, HOT75m3, HOT75m4, HOT75m8, MB0m1, MB0m2, MB20m2, MB20m5, MB20m12, MB40m1, MB40m5, MB40m12, MB100m5, MB100m7, MB100m9, MB100m10, PALB1, PALB2, PALB5, PALB7, PALB6, PALB8, PALE1, PALE6 and PALE7.
95. The light-driven energy generator of claim 88, wherein said host is an artificial membrane system.

96. The light-driven energy generator of claim 88, wherein said host is a cell membrane obtained from a bacterium.
97. The light-driven energy generator of claim 96, wherein said host is a cell membrane preparation obtained from a bacterium.
98. The light-driven energy generator of claim 88, wherein said host is a cell membrane obtained from an eukaryote.
99. The light-driven energy generator of claim 98, wherein said host is a cell membrane preparation obtained from an eukaryote.
100. The light-driven energy generator of claim 88, further comprising a light source for illuminating said light absorbing pigment, whereby said energy generator converts light into biochemical energy.
101. The light-driven energy generator of claim 100, wherein said light source is a fast-pulsed light source.
102. The light-driven energy generator of claim 101, wherein said fast-pulsed light source comprises a mechanism for delivering intermittent fast-light pulses at predetermined time intervals.

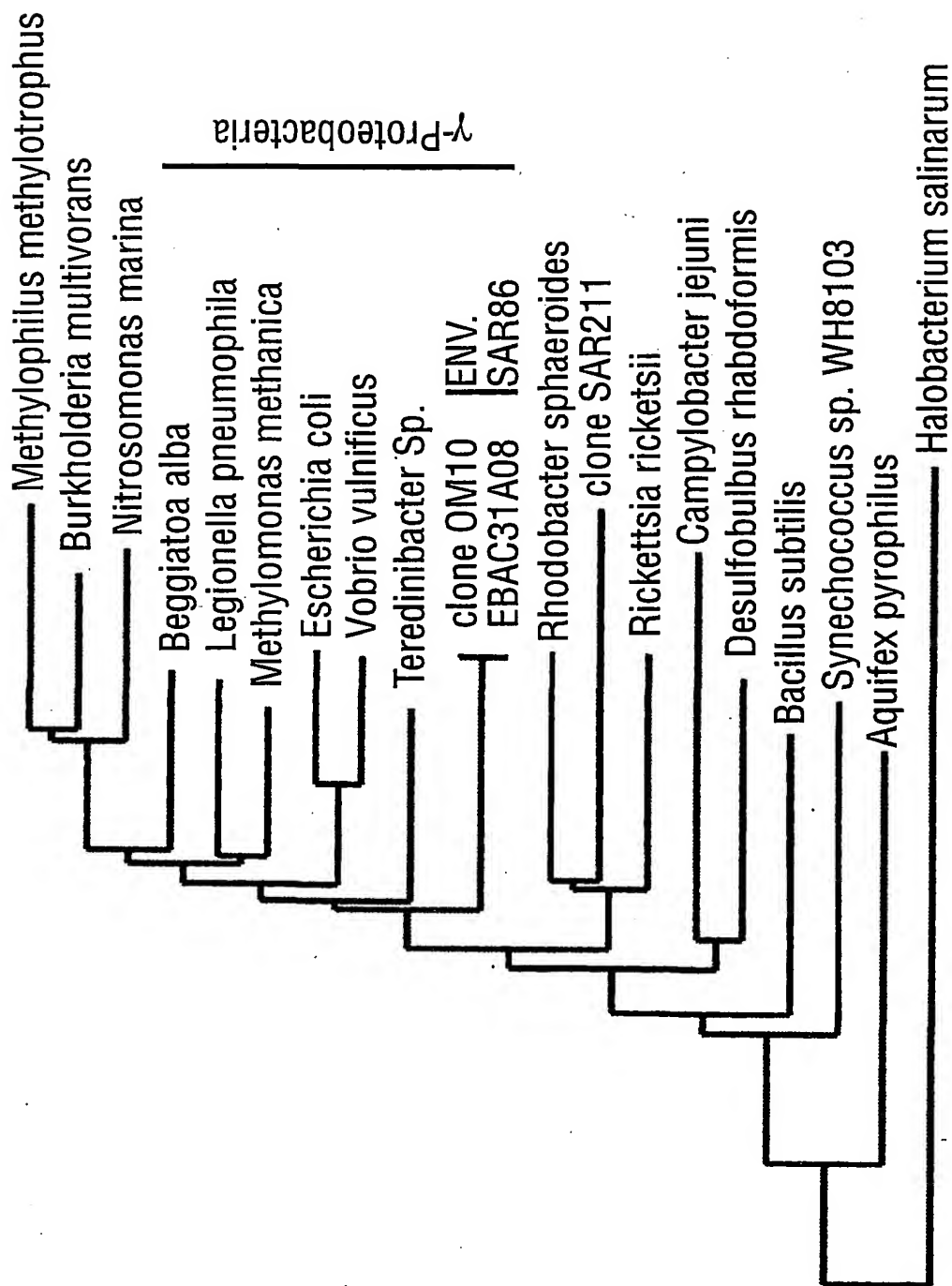
103. The light-driven energy generator of claim 100, wherein said light source is a light source exhibiting different predetermined wavelengths.
104. The light-driven energy generator of claim 88, further comprising a mediator for mediating energy generated by said energy generator into chemical, mechanical or electrical energy.
105. The light-driven energy generator of claim 88, wherein said proteorhodopsin protein is selected to determine an absorption spectra of said light absorbing pigment.
106. A method for making a light-driven energy generator, comprising the steps of:
- (a) providing a proteorhodopsin protein;
 - (b) providing a host to correctly fold said proteorhodopsin protein in said host, thereby creating an integrated proteorhodopsin protein; and
 - (c) providing a source of retinal to bind covalently to said integrated proteorhodopsin protein, thereby creating a light absorbing pigment.
107. The method of claim 106, wherein said proteorhodopsin protein is encoded by a proteorhodopsin gene retrieved from a genomic fragment of a sample of naturally occurring bacteria.
108. The method of claim 107, wherein said naturally occurring bacteria are marine proteobacteria.

109. The method of claim 107, wherein said naturally occurring bacteria are SAR86 bacteria.
110. The method of claim 107, wherein said naturally occurring bacterial genomic fragment is retrieved from a recombinant DNA library.
111. The method of claim 110, wherein said naturally occurring bacterial genomic fragment is retrieved from a bacterial artificial chromosome library.
112. The method of claim 107, wherein said genomic fragment is retrieved from a clone, wherein said clone is a member of the group consisting of BAC31A8, BAC40E8, BAC41B4, BAC64A5, HOT0m1, HOT75m1, HOT75m3, HOT75m4, HOT75m8, MB0m1, MB0m2, MB20m2, MB20m5, MB20m12, MB40m1, MB40m5, MB40m12, MB100m5, MB100m7, MB100m9, MB100m10, PALB1, PALB2, PALB5, PALB7, PALB6, PALB8, PALE1, PALE6 and PALE7.
113. The method of claim 106, wherein said host is an artificial membrane system.
114. The method of claim 106, wherein said host is a cell membrane obtained from a bacterium.

115. The method of claim 114, wherein said host is a cell membrane preparation obtained from a bacterium.
116. The method of claim 106, wherein said host is a cell membrane obtained from an eukaryote.
117. The method of claim 116, wherein said host is a cell membrane preparation obtained from an eukaryote.
118. The method of claim 106, further comprising the step of providing a light source for illuminating said light absorbing pigment, whereby said energy generator converts light into biochemical energy.
119. The method of claim 118, wherein said light source is a fast-pulsed light source.
120. The method of claim 119, wherein said fast-pulsed light source comprises a mechanism for delivering intermittent fast-light pulses at predetermined time intervals.
121. The method of claim 118, wherein said light source is a light source exhibiting different predetermined wavelengths.
122. The method of claim 106, further comprising the step of providing a mediator for mediating energy generated by said energy generator into chemical, mechanical or electrical energy.

123. The method of claim 106, wherein said proteorhodopsin protein is selected to determine an absorption spectra of said light absorbing pigment.
124. A PCR apparatus for amplifying a proteorhodopsin gene from DNA samples of naturally occurring microbial populations using polymerase chain reaction, comprising oligodeoxynucleotide primers with a Watson-Crick base pair complementarity to 5' and 3' ends of said proteorhodopsin gene.
125. The apparatus of claim 124, wherein said primers are according to Sequence ID No:2 and Sequence ID No:3.
126. A method of designing PCR primers, comprising the steps of:
- (a) determining a DNA sequence of a proteorhodopsin gene; and
 - (b) based on said determined DNA sequence in (a), designing oligodeoxynucleotide primers with a Watson-Crick base pair complementarity to said 5' and 3' ends of said proteorhodopsin gene.
127. The method of claim 126, further comprising the step of using said oligodeoxynucleotide primers to amplify said proteorhodopsin gene from DNA samples of naturally occurring microbial populations by polymerase chain reaction.
128. The method of claim 127, further comprising the step of cloning said amplified proteorhodopsin gene into an expression vector.

129. The method of claim 126, wherein said primers are according to Sequence ID No:2 and Sequence ID No:3.

**Fig. 1**

29

accatgggta aattattact gatattagg

Figure 2

24

agcattagaa gattctttaa cagc

Figure 3

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca	48
Met Gly Lys Leu Leu 5	
1	
ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt	96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val	
20 25 30	
tct ttt tgg tta gtt act gct gct tta tta gca tct act gta ttt ttc	144
Ser Phe Trp Leu Val Thr Ala Thr Ala Leu Ala Ser Thr Val Phe Phe	
35 40 45	
ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tta act	192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	
gta tct ggt ctt gtt act ggt att gct ttc ttc tgg cat tac atg tac atg	240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met	
65 70 75 80	
aga ggg gta tgg att gaa act ggt gat tgc cca act gta ttt aga tac	288
Arg Gly Val Trp Ile Glu Thr Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr	
85 90 95	
att gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta	336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu	
100 105 110	

Figure 4

att ctt gct gct gca act aat gtt gct gga tca tta ttt aag aaa tta	384
Ile Leu Ala Ala gct gct gca act Thr Asn Val Ala gct gga tca tta ttt aag aaa tta	
115 120	
cta gtt ggt tct ctt gtt atg ctt gtt ggt ttt ggt tac atg ggt gaa gca	432
Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala	
130 135 140	
gga atc atg gct gca tgg cct gca ttc att att ggg tgt tta gct tgg	480
Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp	
145 150 155 160	
gta tac atg att tat gaa tta tgg gct gga gaa gga aaa tct gca tgt	528
Val Tyr Met Ile Tyr Glu Leu Trp Ala gct gga Gly Glu Gly Lys Ser Ala Cys	
165 170 175	
aat act gca agt cct gct gct gct gct tca cca gct tac aac aca atg atg tat	576
Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr	
180 185 190	
att atc atc ttt ggt tgg gcg att tat cct gta ggt tat ttc aca ggt	624
Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly	
195 200 205	
tac ctg atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat	672
Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr	
210 215 220	

Figure 4

aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720
Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Trp 240
225 230 235

aat gtt gct gtt aaa gaa tct tct aat gct 750
Asn Val Ala Val Lys Glu Ser Ser Asn Ala 250
245

Figure 4

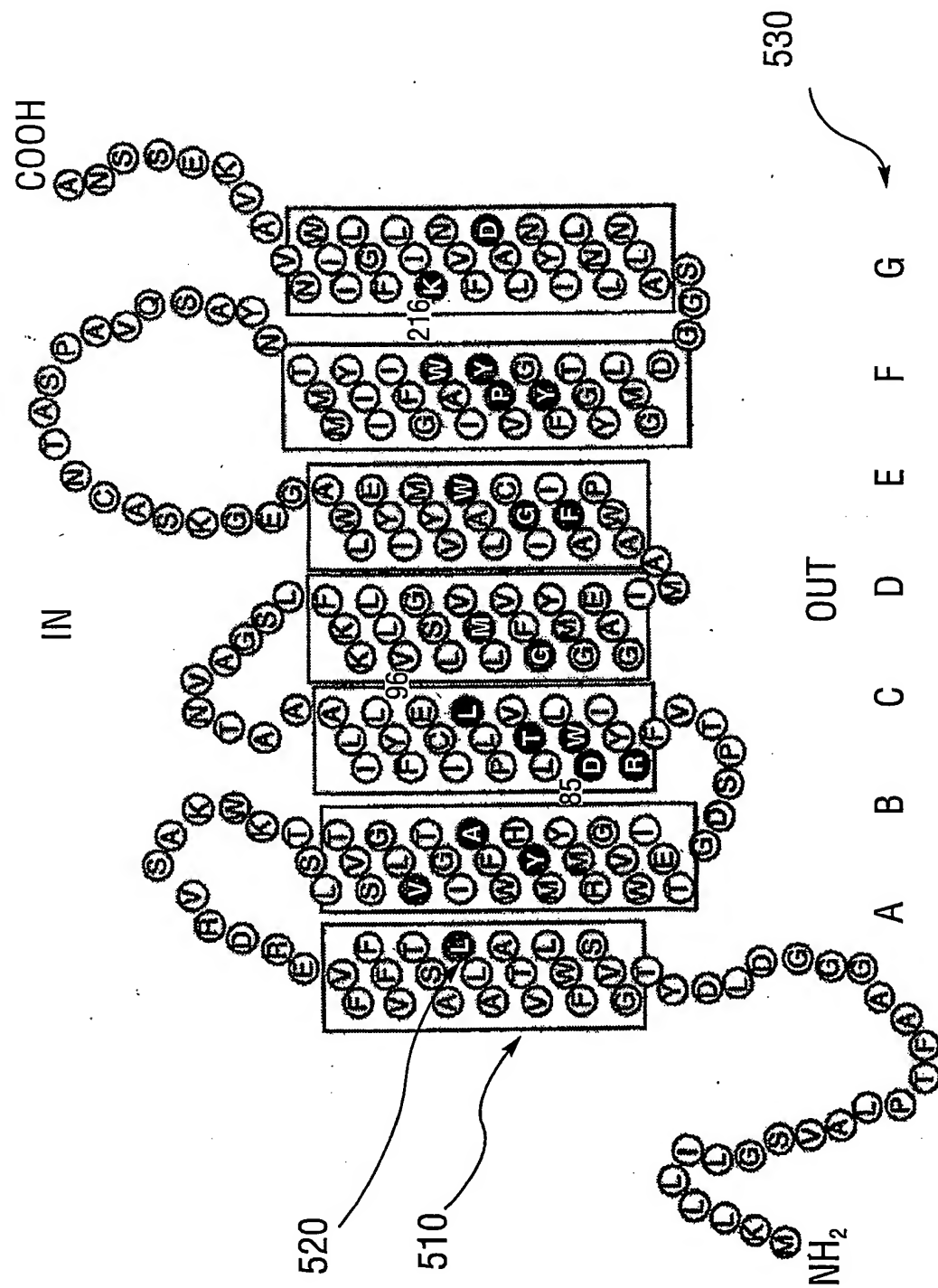


Fig. 5

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca	48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr	
1 5 10 15	
ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt	96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val	
20 25 30	
tct ttt tgg tta gtt act gct gct cta tta gca tct act gta ttt ttc	144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe	
35 40 45	
ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tta act	192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	
gta tcg ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg	240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met	
65 70 75 80	
aga ggg gta tgg att gag act ggt gat tcg cca act gta ttt aga tac	288
Arg Gly Val Trp Ile Glu Thr Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr	
85 90 95	
att gat tgg tta cta aca gtt cct cta ttg ata tgt gaa ttc tac tta	336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu	
100 105 110	

Figure 6

att ctt gct gct gca aca aat gtt gct gct ggc ctg ttt aag aaa tta	384
Ile Leu Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu	
115 120 125	
ttg gtt ggt tct ctt gtt atg ctt gtt ggt tac atg ggt gag gca	432
Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala	
130 135 140	
gga att atg aac gct tgg ggt gca ttc gtt att ggg tgt tta gct tgg	480
Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp	
145 150 155 160	
gta tac atg att tat gaa cta tgg gct gga gaa ggc aag gct gca tgt	528
Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ala Ala Cys	
165 170 175	
aat act gca agt cct gct gtg caa tca gct tac aac aca atg atg tat	576
Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr	
180 185 190	
ata atc atc ttt ggt tgg gca att tat cct gta ggt tat ttc aca ggt	624
Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly	
195 200 205	
tac cta atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat	672
Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr	
210 215 220	

Figure 6

gac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720
Asp Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Trp 240
225 230 235

aat gtt gct gtt aaa gaa tct tct aat gct 750
Asn Val Ala Val Lys Glu Ser Ser Asn Ala 250
245

Figure 6

```

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca      48
Met Gly Lys Leu Leu 5 tta ggt gac ctt gat gct agt gta tta tta tta tta
1 10 15
ttt gct gca ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt
Phe Ala Ala Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
20 25 30
tct ttt tgg tta gct act gct gct gct tta tta gca tct act gta ttt ttc      144
Ser Phe Trp Leu Ala Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45
ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act      192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60
gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg      240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80
aga ggg gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac      288
Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85 90 95
att gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta      336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

```

Figure 7

aac	ctt	gct	gat	ttt	gtt	aac	aag	att	cta	ttt	ggt	tta	att	ata	tgg	720
Asn	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile	Trp	
225					230					235					240	
aat	gtt	gct	gtt	aaa	gaa	tct	tct	aat	gct							750
Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala							
				245					250							

Figure 7

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca	48
Met Gly Lys Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr	
1 5 10 15	
ttt gct gca ggt ggc ggt gac ctt gat gct agt gat tac act ggt gtt	96
Phe Ala Ala Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val	
20 25 30	
tct ttt tgg tta gtt aca gct gct cta tta gca tct act gta ttt ttc	144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe	
35 40 45	
ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tta act	192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	
gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg	240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met	
65 70 75 80	
aga gga gta tgg att gaa act ggt gat tcg cct act act gta ttt aga tac	288
Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr	
85 90 95	
att gat tgg tta cta aca gtt cct tta tta ata tgt gaa ttc tac tta	336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu	
100 105 110	

Figure 8

```

att ctt gct gct gca act aat gtt gcc ggc tca tta ttt aag aaa ctt      384
Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
115 120
cta gtt ggt tct ctt gtt atg ctt gtt ggt ttt ggt tac atg ggt gaa gca      432
Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140
gga att atg gca gct tgg cct gca ttc att att ggg tgt tta gct tgg      480
Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
145 150 155 160
gta tac atg att tat gaa cta tat gct gga gaa gga aaa tct gca tgt      528
Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys
165 170
aat act gca agt cct tcg gtt caa tca gct tac aac aca atg atg gct      576
Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala
180 185 190
atc ata gtc ttc ggt tgg gca att tat cct ata ggt tat ttc aca ggt      624
Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Ile Gly Tyr Phe Thr Gly
195 200 205
tac cta atg ggt gac ggt gga tca gct ctt aac tta aac ctt att tat      672
Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr
210 215 220

```

Figure 8

aac	ctt	gct	gac	ttt	gtt	aac	aag	att	cta	ttt	ggt	tta	att	ata	tgg	720
Asn	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile	Trp	
225					230					235					240	
aat	gtt	gct	gtt	aaa	gaa	tct	tct	aat	gct							750
Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala							
				245					250							

Figure 8

EBAC31A8	1	atgggtaaat	tattactgat	attaggTAGT	GTTATTGCAC	TTCC	TACATTT	50
EBAC40	1	50
EBAC41	1	50
EBAC64	1	50
EBAC31A8	51	TGCTGCAGGT	GGTGGTGACC	TTGATGCTAG	TGATTACACT	GGTGT	TTTCTT	100
EBAC40	51	100
EBAC41	51	100
EBAC64	51	100
EBAC31A8	101	TTTGGTTAGT	TACTGCTGCT	TTATTAGCAT	CTACTGTATT	TTTCT	TTTGT	150
EBAC40	101	150
EBAC41	101	150
EBAC64	101	150
EBAC31A8	151	GAAAGAGATA	GAGTTTCTGC	AAAATGGAAA	ACATCATTAA	CTGTAT	CTCTGG	200
EBAC40	151	200
EBAC41	151	200
EBAC64	151	200
EBAC31A8	201	TCCTTGTTACT	GGTATTGCTT	TCTGGCATT	CATGTACATG	AGAGGGG	TAT	250
EBAC40	201	250
EBAC41	201	250
EBAC64	201	250
EBAC31A8	251	GGATTGAAAC	TGGTGATTG	CCAAC	TGTAT	TGATTGG	TTA	300
EBAC40	251	300
EBAC41	251	300
EBAC64	251	300

Figure 9

EBAC31A8	301	CTAACAGTTC	CTCTATTAAAT	ATGTGAATTC	TACTTAATTC	TTGCTGCTGC	350
EBAC40	301G.	350
EBAC41	301	350
EBAC64	301T.	350
EBAC31A8	351	AACTAATGTT	GCTGGATCAT	TATTTAAGAA	ATTACTAGTT	GGTTCTCTTG	400
EBAC40	351	..A.....	...CTGGCC	.G.....	...T.G...	400
EBAC41	351	T.....	400
EBAC64	351C.C....C.T.	400
EBAC31A8	401	TTATGCTTGT	GTTTGGTTAC	ATGGGTGAAG	CAGGAATCAT	GGCTGCATGG	450
EBAC40	401G.T..	..AAC..T...	450
EBAC41	401	450
EBAC64	401T..	...A..T...	450
EBAC31A8	451	CCTGCATTCA	TTATTGGGTG	TTTAGCTTGG	GTATACATGA	TTTATGAATT	500
EBAC40	451	GG.....GC.	500
EBAC41	451C.	500
EBAC64	451C.	500
EBAC31A8	501	ATGGGCTGGA	GAAGGAAAAT	CTGCATGTAA	TACTGCAAGT	CCTGCTGTGC	550
EBAC40	501C..GG	550
EBAC41	501	550
EBAC64	501	..AT.....T.G..T.	550
EBAC31A8	551	AATCAGCTTA	CAACACAAATG	ATGTATATTA	TCATCTTTGG	TTGGGCGATT	600
EBAC40	551A.A..	600
EBAC41	551	600
EBAC64	551GC...C.	..AG....C..A...	600

Figure 9

EBAC31A8	601	TATCCTGTAG	GTTATTTCAC	AGGTTACCTG	ATGGGTGACG	GTGGATCAGC	650
EBAC40	601A	650
EBAC41	601	650
EBAC64	601AA	650
EBAC31A8	651	TCTTAACTTA	AACCTTATCT	ATAACCTTGC	TGACTTTTGT	AACAAGATTC	700
EBAC40	651G	700
EBAC41	651T	700
EBAC64	651T	700
EBAC31A8	701	TATTTGTTT	AATTATATGG	AATGTTgctg	ttaaagaatc	ttctaagtct	750
EBAC40	701	750
EBAC41	701	750
EBAC64	701	750

Figure 9

EBAC31A8	1	MGKLLILIGS	VIALPTFAAG	GGDLASDYT	GVSEWLVTA	LLASTVFEFV	50
EBAC40_1	1	50
EBAC41_1	1A.....	50
EBAC64_1	1	50
EBAC31A8	51	ERDRVSAKWK	TSLTVSGLVT	GIAFWHYMYM	RGVWIETGDS	PTVFRYIDWL	100
EBAC40_1	51	100
EBAC41_1	51	100
EBAC64_1	51	100
EBAC31A8	101	LTVPLLICEF	YLILAAATNV	AGSLFKKLLV	GSLVMLVFGY	MGEAGIMAAW	150
EBAC40_1	101AG.....N.....	150
EBAC41_1	101	150
EBAC64_1	101	150
EBAC31A8	151	PAFIGCLAW	VYMIYELWAG	EGKSACNTAS	PAVQSAYNMT	MYIIIFGWAI	200
EBAC40_1	151	G..V.....A.....	200
EBAC41_1	151	200
EBAC64_1	151Y.....S.....A..V.....	200
EBAC31A8	201	YPVGYFTGYL	MGDGGGSA	NLIYNLADFV	NKILFGLIIW	NVAVKESSNA	250
EBAC40_1	201D.....	250
EBAC41_1	201	250
EBAC64_1	201	..I.....	250

Figure 10

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca
 Met Gly Lys Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15 48

 ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt
 Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30 96

 tct ttt tgg tta gtt act gct gct gct cta tta gca tct act gta ttt ttc
 Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45 144

 ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act
 Phe Val Glu Arg Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60 192

 gta tcg ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80 240

 aga ggg gta tgg att gag acc ggt gat tcg cca act gta ttt aga tac
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95 288

 att gat tgg tta cta aca gtt cct cta ttg ata tgt gaa ttc tac tta
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
 100 105 110 336

Figure 11

att ctt gct gct gca aca aat gtt gct gct ggc ctg ttt aag aaa tta 384
 Ile Leu Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu
 115 120 125
 ttg gtt ggt tct ctt gtt atg ctt gtt ggt tac atg ggt gag gca 432
 Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
 130 135 140
 gga att atg aac gct tgg ggt gca ttc gtt att ggg tgt tta gct tgg 480
 Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp
 145 150 155 160
 gta tac atg att tat gaa cta tgg gct gga gaa ggc aag gct gca tgt 528
 Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ala Ala Cys
 165 170 175
 aat act gca agt cct gct gtg caa tca gct tac aac aca atg atg tat 576
 Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
 180 185 190
 ata atc atc ttt ggt tgg gca att tat cct gta ggt tat ttc aca ggt 624
 Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205
 tac cta atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat 672
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr
 210 215 220

Figure 11

aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720
Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Trp 240
225 230 235

aat gtt gct gtt aaa gaa tct tct aat gct 750
Asn Val Ala Val Lys Glu Ser Ser Asn Ala 250
245 250

Figure 11

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
 Met Gly Lys Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15
 ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
 Phe Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30
 tca ttc tgg ctg ctg gtt aca gct gct ggt atg tta gcg gca act gtg ttc ttt 144
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Thr Val Phe Phe
 35 40 45
 ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt gct 192
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Ala
 50 55 60
 gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg 240
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80
 aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat 288
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95
 att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
 100 105 110

Figure 12

att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt 384
 Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
 115 120 125

 cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct 432
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
 130 135 140

 gga tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg 480
 Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
 145 150 155 160

 tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta 528
 Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
 165 170 175

 agt act gca agt cct gct gct gtt aac tct gca tac aac gca atg atg atg 576
 Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met
 180 185 190

 att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt 624
 Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
 195 200 205

 tac cta atg ggt ggc gaa ggt gta tac gct tca aac tta aac ctt ata 672
 Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
 210 215 220

Figure 12

tat aac ctt gct gcc gac ctt gtt aac aag att cta ttt ggt ttg atc att 720
Tyr Asn Leu Ala Asp Leu Val Asn Lys Ile Leu Phe Gly Leu Ile Ile 240
225 230 235

tgg aat gtt gct gtt aaa gaa tct tct aat gct 753
Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala 250
245

Figure 12

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15
 ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
 Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30
 tca ttc tgg ctg ctg gtt aca gct ggt atg tta gcg gca act gta ttc ttt 144
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Thr Val Phe Phe
 35 40 45
 ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act 192
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60
 gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tac atg 240
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80
 aga ggt gtt tgg ata gat act ggt gat aca cca aca gta ttt aga tat 288
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95
 att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
 100 105 110

Figure 13

att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt 384
 Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
 115 120 125

 cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct 432
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
 130 135 140

 ggt tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg 480
 Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
 145 150 155 160

 tta tac atg att tat gag cta cat atg ggt gaa ggt aag gct gct gta 528
 Leu Tyr Met Ile Tyr Glu Leu His Met Gly Glu Tyr Lys Ala Ala Val
 165 170 175

 agt act gca agt cct gct gct gtt aac tct gca tac aac gca atg atg aag 576
 Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Lys
 180 185 190

 att att gtt att gga tgg gca att tat cct gct gga tat gct gct ggt 624
 Ile Ile Val Ile Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
 195 200 205

 tac cta atg agt ggt gac ggt gta tac gct tca aac tta aac ctt ata 672
 Tyr Leu Met Ser Gly Asp Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
 210 215 220

Figure 13

tat aac ctt gct gac gtt gtt aac aag att cta ttt ggt ttg atc att	720
Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile	240
225	230
235	240
250	255
260	265
270	275
280	285
290	295
300	305
310	315
320	325
330	335
340	345
350	355
360	365
370	375
380	385
390	395
400	405
410	415
420	425
430	435
440	445
450	455
460	465
470	475
480	485
490	495
500	505
510	515
520	525
530	535
540	545
550	555
560	565
570	575
580	585
590	595
600	605
610	615
620	625
630	635
640	645
650	655
660	665
670	675
680	685
690	695
700	705
710	715
720	725
730	735
740	745
750	755
760	765
770	775
780	785
790	795
800	805
810	815
820	825
830	835
840	845
850	855
860	865
870	875
880	885
890	895
900	905
910	915
920	925
930	935
940	945
950	955
960	965
970	975
980	985
990	995
1000	1005

Figure 13

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
 Met Gly Lys Leu Leu 5
 1 10 15
 ttt gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
 Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30
 tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gtg ttc ttt 144
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
 35 40 45
 ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act 192
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60
 gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg 240
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80
 aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat 288
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95
 att gat tgg tta tta act gtt cca tta caa gtg gtt gag ttc tat cta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Val Val Glu Phe Tyr Leu
 100 105 110

Figure 14

att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt 384
 Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
 115 120 125

 cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct 432
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
 130 135 140

 gga tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg 480
 Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
 145 150 155 160

 tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta 528
 Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
 165 170 175

 agt act gca agt cct gct gct gtt aac tct gca tac aac gca atg atg atg 576
 Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met
 180 185 190

 att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt 624
 Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
 195 200 205

 tac cta atg ggt ggc gaa ggt gta tac gct tca aac tta aac ctt ata 672
 Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
 210 215 220

Figure 14

tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att	720
Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile	240
225	230
tgg aat gtt gct gtt gaa gaa tct tct aat gct	753
Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala	250
245	240

Figure 14

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15
 ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
 Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30
 tca ttc tgg ctg gtt gtt aca gct ggt atg tta gcg gca act gtg ttc ttt 144
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Thr Val Phe Phe
 35 40 45
 ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act 192
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60
 gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg 240
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80
 aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat 288
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95
 att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
 100 105 110

Figure 15

att ctt gct gct tgt aca aat gtt gct gct tca tta ttt aag aag ctt 384
 Ile Leu Ala Ala Cys Thr Asn Val Ala Ala Ser Leu Phe Lys Lys Leu
 115 120 125
 cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct 432
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
 130 135 140
 gga ttg gct cct gta tgg cct gct ttc att att ggt atg gct gga tgg 480
 Gly Leu Ala Pro Val Trp Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
 145 150 155 160
 tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta 528
 Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
 165 170 175
 agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg gtg 576
 Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Val
 180 185 190
 att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt 624
 Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
 195 200 205
 tac cta atg ggt ggc gaa ggt gta tac gct gct tca aac tta aac ctt ata 672
 Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
 210 215 220

Figure 15

tat aac ctt gcc gac gtt gtt aac aag att cta ttt ggt ttg atc att 720
Tyr Asn Leu Ala Asp Leu Val Asn Lys Ile Leu Phe Gly Leu Ile Ile 240
225 230 235

tgg aat gtt gct gtt gaa gaa tct tct aat gct 753
Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala 250
245

Figure 15

```

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca      48
Met Gly Lys Leu Leu 5 tta ctg ata tta ggt agt gtt att gca ctt cct aca
1                               10
                               15
tth gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt      96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20                               25
                               30
tct ttt tgg tta gtt act gct gct gct cta tta gca tct act gta ttt ttc      144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35                               40
                               45
tth gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act      192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50                               55
                               60
gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg      240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65                               70
                               75
                               80
aga ggg gta tgg att gag act ggt gat gat tgg cca act gta ttt aga tac      288
Arg Gly Val Trp Ile Glu Thr Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85                               90
                               95
att gat tgg tta cta aca gtt cct cta ttg ata tgt gaa ttc tac tta      336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100                               105
                               110

```

Figure 16

att ctt gct gct gca aca aat gtt gct gct ggc ctg ttt aag aaa tta 384
 Ile Leu Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu
 115 120 125
 ttg gtt ggt tct ctt gtt atg ctt gtt ggt tac atg ggt gag gca 432
 Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
 130 135 140
 gga att atg aac gct tgg cct gca ttc att att ggt tta gct tgg 480
 Gly Ile Met Asn Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
 145 150 155 160
 gta tac atg att tat gaa cta tat gct gga gaa gga aaa tct gca tgt 528
 Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Tyr Ser Ala Cys
 165 170 175
 aat act gca agt cct tcg gtt caa tca gct tac aac aca atg atg gct 576
 Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala
 180 185 190
 atc ata gtc ttc ggt tgg gca att tat cct gta ggt tat ttc aca ggt 624
 Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205
 tac cta atg ggt gac ggt gga tca gct ctt aac tta aac ctt att tat 672
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr
 210 215 220

Figure 16

aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720
Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Trp 240
225 230 235

aat gtt gct gtt aaa gaa tct tct aat gct 750
Asn Val Ala Val Lys Glu Ser Ser Asn Ala 250
245 250

Figure 16

```

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca      48
Met Gly Lys Leu Leu 5 tta ctg ata tta ggt agt gtt att gca ctt cct aca
1                               10
tct gct gca ggt ggt ggt gac ctt gat gct agt gat gtt ggt ggt gtt      96
Phe Ala Ala Gly Gly Gly Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
20                               25
tct ttt tgg tta gtt act gct gct gct tta tta gca tct act gta ttt ttc      144
Ser Phe Trp Leu Val Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
35                               40
tct gtt gaa aga gat aga gtt tct tct gca aaa tgg aaa aca tta act      192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50                               60
gta tct ggt ctt gtt act ggt act ggt att gct ttc tgg cat tac atg tac atg      240
Val Ser Gly Leu Val Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
65                               70
aga ggg gta tgg att gaa act ggt gat tgg cca act gta ttt aga tac      288
Arg Gly Val Trp Ile Glu Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
85                               90
att gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta      336
Ile Asp Trp Leu Leu Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
100                               110

```

Figure 17

Figure 17

Figure 17

aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720
Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Trp 240
225 230

aat gtt gct gtt aaa gaa tct tct aat gct 750
Asn Val Ala Val Lys Glu Ser Ser Asn Ala 250
245

Figure 17


```

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca      48
Met Gly Lys Leu Leu Ile Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1      5      10      15

ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt ggt      96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20      25      30

tct ttt tgg tta gtt act gct gct tta tta gca tct act gta ttt ttc      144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35      40      45

ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tta act      192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50      55      60

gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg      240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65      70      75      80

aga ggg gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac      288
Arg Gly Val Trp Ile Glu Thr Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85      90      95

att gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta      336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100      105      110

```

Figure 18

Figure 18

aac	ctt	gct	gac	ttt	gtt	aac	aag	att	cta	ttt	ggt	tta	att	ata	tgg	720
Asn	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile	Trp	
225					230					235					240	
aat	gtt	gct	gtt	aaa	gaa	tct	tct	aat	gct							750
Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala							
				245					250							

Figure 18

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca 48
 Met Gly Lys Leu Leu 5
 1
 ttt gct gca ggt ggc ggt gac ctt gat gct agt gat tac act ggt gtt 96
 Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30
 tct ttt tgg tta gtt aca gct gct cta tta gca tct act gta ttt ttc 144
 Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45
 ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tta act 192
 Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60
 gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg 240
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80
 aga ggg gta tgg att gaa act ggt gat tgc cca act gta ttt aga tac 288
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95
 att gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
 100 105 110

Figure 19

aac ctt gct gac ttt gtt aac aag att cta ctt ggt tta att ata tgg 720
Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Leu Gly Leu Ile Trp 240
225 230

aat gtt gct gtt aaa gaa tct tct aat gct 750
Asn Val Ala Val Lys Glu Ser Ser Asn Ala 250
245

Figure 19

48
 atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15
 96
 ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt
 Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30
 144
 tct ttt tgg tta gtt act gct gct tta gca tct act gta ttt ttc
 Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45
 192
 ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tta act
 Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60
 240
 gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80
 288
 aga ggg gta tgg att gaa act ggt gat tgg cca act gta ttt aga tac
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95
 336
 att gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
 100 105 110

Figure 20

att ctt gct gct gca gct aat gtt gct gga tca tta ttt aag aaa tta	384
Ile Leu Ala Ala Ala Val Asn Val Ala Gly Ser Leu Phe Lys Lys Leu	
115 120 125	
cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca	432
Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala	
130 135 140	
gga atc atg gct gca tgg cct gca ttc att att ggg tgt tta gct tgg	480
Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp	
145 150 155 160	
gta tac atg att tat gaa tta tgg gct gga gaa gga aaa tct gca tgt	528
Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys	
165 170 175	
aat act gca agt cct gct gct gtg caa tca gcc tac aac aca atg atg tat	576
Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr	
180 185 190	
att atc atc ttt ggt ggt tgg gcg att tat cct gta ggt tat ttc aca ggt	624
Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly	
195 200 205	
tac ttg atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat	672
Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr	
210 215 220	

Figure 20

aac	ctt	gct	gac	ttt	ggt	aac	aag	att	cta	ttt	ggt	tta	att	ata	tgg	720
Asn	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile	Trp	
225					230					235					240	
aat	ggt	gct	ggt	aaa	gaa	tct	tct	aat	gct							750
Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala							
				245					250							

Figure 20

atg ggt aaa tta tta ctg ata ata ggt agt gtt att gca ctt cct aca 48
 Met Gly Lys Leu Leu 5
 1
 ttt gct gca ggt ggc ggt gac ctt gat gct agt gat tac act ggt gtt 96
 Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30
 tct ttt tgg tta gtt aca gct gct gct cta tta gca tct act gta ttt ttc 144
 Ser Phe Trp Leu Val Thr Ala Thr Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45
 ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192
 Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60
 gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg 240
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met 80
 65 70 75
 aga gga gta tgg att gaa act ggt gat gct tgc cca act gta ttt aga tac 288
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95
 att gat tgg tta cta aca gtt cct tta ata tgt gaa ttc tac tta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
 100 105 110

Figure 21

att ctt gct gct gca act aat gtt gcc ggc tca tta ttt aag aaa ctt 384
 Ile Leu Ala Ala 115 Val Asn 120 Ala Gly Ser Leu Phe 125 Lys Lys Leu

 cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca 432
 Leu Val Gly Ser Leu Val Met 135 Leu Val Phe Gly Tyr Met Gly Glu Ala
 130

 gga att atg gca gct tgg cct gca ttc att att ggg tgt tta gct tgg 480
 Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
 145 150 155 160

 gta tat atg att tat gaa cta tat gct gga gaa gga aaa tct gca tgt 528
 Val Tyr Met Ile Tyr 165 Tyr Ala Gly Glu Gly Lys Ser Ala Cys
 170 175

 aat aca gca agt cct gct gtg caa tca gct tac aac aca atg atg tat 576
 Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
 180 185 190

 att atc gtc ttt ggt tgg gcg att tat cct gta ggt tat ttc aca ggt 624
 Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205

 tac ctg atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat 672
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr
 210 215 220

Figure 21

aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720
Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Trp 240
225 230 235

aat gtt gct gtt aaa gaa tct tct aat gct 750
Asn Val Ala Val Lys Glu Ser Ser Asn Ala 250
245

Figure 21

```

atg ggt aaa tta tta tta ctg ata tta ggt agt gtt att gca ctt cct aca      48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1      5      10      15

ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt      96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20      25      30

tct ttt tgg tta gtt act gct gct gct cta tta gca tct act gta ttt ttc      144
Ser Phe Trp Leu Val Thr Ala Thr Ala Leu Ala Ser Thr Val Phe Phe
35      40      45

ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act      192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50      55      60

gta tcg ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg      240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65      70      75      80

aga ggg gta tgg att gag act gag act ggt gat tcg cca act gta ttt aga tac      288
Arg Gly Val Trp Ile Glu Thr Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85      90      95

att gat tgg tta cta aca gtt cct cta ttg ata tgt gaa ttc tac tta      336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100      105      110

```

Figure 22

att ctt gct gct gca aca aat gtt gct gct ggc ctg ttt aag aaa tta 384
 Ile Leu Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu
 115 120 125
 ttg gtt ggt tct ctt gtt atg ctt gtt ggt tac atg ggt gag gca 432
 Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
 130 135 140
 gga att atg aac gct tgg ggt gca ttc gtt att ggg tgt tta gct tgg 480
 Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp
 145 150 155 160
 gta tac atg att tat gaa cta tgg gct gga gaa ggc aag gct gca tgt 528
 Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ala Ala Cys
 165 170 175
 aat act gca agt cct gct gtg caa tca gct tac aac aca atg atg tat 576
 Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
 180 185 190
 ata atc atc ttt ggt gca att tat cct gta ggt tat ttc aca ggt 624
 Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205
 tac cta atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat 672
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr
 210 215 220

Figure 22

aac ctt gct gac ttt gtt aac aag aat cta ttt ggt tta att ata tgg 720
Asn Leu Ala Asp Phe Val Asn Lys Asn Leu Phe Gly Leu Ile Ile Trp 240
225 230 235

aat gtt gct gtt aaa gaa tct tct aat gct 750
Asn Val Ala Val Lys Glu Ser Ser Asn Ala 250
245

Figure 22

atg ggt aaa tta tta tta cgg ata tta ggt agt gtt att gca ctt cct aca 48
 Met Gly Lys Leu Leu Leu Arg Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15
 ttt gct gca ggt ggc ggt ggt gac ctt gat gct agt gat tac act ggt gtt 96
 Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30
 tct ttt tgg tta gtt aca gct gct gct cta tta gca tct act gta ttt ttc 144
 Ser Phe Trp Leu Val Thr Ala Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45
 ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192
 Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60
 gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tat atg 240
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80
 aga gga gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac 288
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95
 att gat tgg tta cta aca gtt cct tta tta ata tgt gaa ttc tac tta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
 100 105 110

Figure 23

att ctt gct gct gca gct gtt gct gga tca tta ttt aag aaa tta 384
 Ile Leu 115 Ala Ala Ala Thr Asn Val 120 gtt Val Ala Gly Ser Leu Phe Lys Lys Leu 125

 cta gtt ggt tct ctt gtt gtt gtt gtt gtt gtt ggt tac atg ggt gaa gca 432
 Leu Val 130 Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala 140
 135

 gga atc atg gct gca tgg cct gca ttc att att ggg tgt tta gct tgg 480
 Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp 155
 145 150

 gta tac atg att tat gaa cta tgg gct gga gaa gga aaa tct gca tgt 528
 Val Tyr Met Ile Tyr 165 Glu Trp Ala Gly Glu Gly Lys Ser Ala Cys 170
 165

 aat act gca agt cct gct gct gct gct tca gct tac aac aca atg atg tat 576
 Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr 180
 185

 atc atc atc gtt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt 624
 Ile Ile Ile Val Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly 190
 195 200

 tac ctg atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat 672
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr 210
 215 220

Figure 23

720
aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg
Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp 240
225 230
aat gtt gct gtt gaa tct tct aat gct
Asn Val Ala Val Lys Glu Ser Ser Asn Ala 250
245

Figure 23

```

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca      48
Met Gly Lys Leu Leu 5      15
1
ttt gct gca ggt ggc ggt gac ctt gat gct agt gat tac act ggt gtt      96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20      25      30
tct ttt tgg tta gtt aca gct gct gct tta gca tct act gta ttt ttc      144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35      40      45
ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act      192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50      55      60
gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg      240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65      70      75      80
aga gga gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac      288
Arg Gly Val Trp Ile Glu Thr Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85      90      95
att gat tgg tta cta aca gtt cct tta tta ata tgt gaa ttc tac tta      336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100      105      110

```

Figure 24

att ctt gct gct gca act aat gtt gcc ggc tca tta ttt aag aaa ctt 384
 Ile Leu Ala Ala gct gca act Thr Asn Val Ala Gly Ser Leu Phe Lys Leu
 115 120 125

 cta gtt ggt tct ctt gtt atg ctt gtt ggt tac atg ggt gaa gca 432
 Leu Val Gly Ser Leu Leu Val Met Leu Phe Gly Tyr Met Gly Glu Ala
 130 135 140

 gga att atg gca gct tgg cct gca ttc att att ggg tgt tta gct tgg 480
 Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
 145 150 155 160

 gta tac atg att tat gaa cta tat gct gga gaa gga aaa tct gca tgt 528
 Val Tyr Met Ile Tyr Ala Trp Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys
 165 170 175

 aat act gca agt cct tcg gtt caa tca gct gct tac aac aca atg atg gct 576
 Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala
 180 185 190

 atc ata gtc ttc ggt tgg gca att tat cct gta ggt tat ttc aca ggt 624
 Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205

 tac cta atg ggt gac ggt gga tca gct ctt aac tta aac ctt att tat 672
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr
 210 215 220

Figure 24

720
aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg
Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp 240
225 230
aat gtt gct gtt aaa gaa tct tct aat gct
Asn Val Ala Val Lys Glu Ser Ser Asn Ala 250
245

750

Figure 24

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca 48
 Met Gly Lys Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15
 ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt 96
 Phe Ala Ala Gly Gly Gly Asp Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30
 tct ttt tgg tta gtt act gct gct tta tta gca tct act gta ttt ttc 144
 Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45
 ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192
 Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60
 gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg 240
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80
 aga ggg gta tgg att gaa act ggt gat tgg cca act gta ttt aga tac 288
 Arg Gly Val Trp Ile Glu Thr Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95
 att gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
 100 105 110

Figure 25

Figure 25

aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720
Asn Leu Ala Asp Phe Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Trp 240
225 230 235

aat gct gct gtt gtt aaa gaa tct tct aat gct 750
Asn Ala Ala Val Lys Glu Ser Ser Asn Ala 250
245

Figure 25

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca 48
 Met Gly Lys Leu Leu 5
 1
 ttt gct gca ggt ggt ggt gac gct gct gct agt ggt ggt ggt gtt 96
 Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30
 tct ttt tgg tta gtt act gct gct gct tta tta gca tct act gta ttt ttc 144
 Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45
 ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192
 Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60
 gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tat atg 240
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80
 aga ggg gta tgg att gaa act ggt ggt gat tgg cca act gta ttt aga tac 288
 Arg Gly Val Trp Ile Glu Thr Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95
 ata gat tgg tta cta aca gtt cct tta tta ata tgt gaa ttc tac tta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
 100 105 110

Figure 26

Figure 26

720
aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg
Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Trp 240
225 230 235
aat gtt gct gtt gaa tct tct aat gct
Asn Val Ala Val Lys Glu Ser Ser Asn Ala 250
245

Figure 26

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca 48
 Met Gly Lys Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15
 ttt gct gca ggt ggc ggt gac ctt gat gct agt gat act ggt gtt 96
 Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30
 tct ttt tgg tta gtt aca gct gct gct cta tta gcg tct act gta ttt ttc 144
 Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45
 ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192
 Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60
 gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tat atg 240
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80
 aga gga gta tgg att gaa act ggt gat tgg cca act gta ttt aga tac 288
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95
 att gat tgg tta cta aca gtt cct tta tta ata tgt gaa ttc tac tta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
 100 105 110

Figure 27

att ctt gct gct gca act aat gtt gcc ggc tca tta ttt aag aaa ctt 384
 Ile Leu Ala Ala gct gca act Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
 115 120 125

 cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca 432
 Leu Val Gly Ser Leu Val Met Leu Val Phe Phe Gly Tyr Met Gly Glu Ala
 130 135 140

 gga ata atg gcg gct tgg cct gca ttc atc gtt gga tgt tta gca tgg 480
 Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Val Gly Cys Leu Ala Trp
 145 150 155 160

 gta tat atg att tat gaa cta tgg gct ggt gaa gga aaa tct gca tgt 528
 Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys
 165 170 175

 aat act gca agt cct gct gta cag tca gct tac aac aca atg atg tat 576
 Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
 180 185 190

 atc atc atc gtt ggt gca att tat cct gta ggt tat ttc aca ggt 624
 Ile Ile Ile Val Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205

 tac cta atg ggt gac ggt gga tca gct ctt aat cta aac ctt att tat 672
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr
 210 215 220

Figure 27

aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720
Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Trp 240
225 230 235

aat gtt gct gtt aaa gaa tct tct aat gct 750
Asn Val Ala Val Lys Glu Ser Ser Asn Ala 250
245 250

Figure 27

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
 Met Gly Lys Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15
 ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
 Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30
 tca ttc tgg ctg gtt aca gct gct ggt atg tta gcg gca act gta ttc ttt 144
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Thr Val Phe Phe
 35 40 45
 ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act 192
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60
 gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tac atg 240
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80
 aga ggt gtt tgg ata gat act ggt gat aca cca aca gta ttt aga tat 288
 Arg Gly Val Trp Ile Asp Thr Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95
 att gat tgg cta tta act gtt cca tta caa atg gtt gag ttc tat cta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
 100 105 110

Figure 28

att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt	384
Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu	
115 120 125	
cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct	432
Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala	
130 135 140	
ggt tta gct cct gta tta cct gct ttc att ctt ggt atg gct ggt tgg	480
Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Leu Gly Met Ala Gly Trp	
145 150 155 160	
tta tac atg att tat gag cta cat atg ggt gaa ggt aag gct gct gta	528
Leu Tyr Met Ile Tyr Glu Leu His Met Gly Glu Lys Ala Ala Val	
165 170 175	
agt act gca agt cct gct gtt aac tct gct tac aat gca atg atg aag	576
Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Lys	
180 185 190	
att att gtt att gga tgg gca att tat cct gct gga tat gct gct ggt	624
Ile Ile Val Ile Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly	
195 200 205	
tac cta atg agt ggt gac ggt gta tac gct tca aac tta aac ctt ata	672
Tyr Leu Met Ser Gly Asp Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile	
210 215 220	

Figure 28

tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att 720
Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile 240
225 230 235

tgg aat gtt gct gtt gaa tct tct aat gct 753
Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala 250
245

Figure 28

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
 Met Gly Lys Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15
 ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
 Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30
 tca ttc tgg ctg gtt aca gct gct ggt atg tta gcg gca act gtg ttc ttt 144
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Thr Val Phe Phe
 35 40 45
 ttt gta gaa aga gac caa gtc agc gct gag tgg aaa act tca ctt act 192
 Phe Val Glu Arg Asp Gln Val Ser Ala Glu Trp Lys Thr Ser Leu Thr
 50 55 60
 gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg 240
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80
 aga ggt gtt tgg ata gat act ggt gat acc cca aca gta ttc aga tat 288
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95
 att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
 100 105 110

Figure 29

att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt	384
Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu	
115 120 125	
cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct	432
Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Glu Ala	
130 135 140	
gga tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg	480
Gly Leu Ala Pro Val Leu Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp	
145 150 155 160	
tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta	528
Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val	
165 170 175	
agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg atg	576
Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met	
180 185 190	
att att gtt gtt gga tgg gca att tat cct gct gct tat gct gct ggt	624
Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly	
195 200 205	
tac cta atg ggt ggc gaa ggt gta tac gct gct tca aac tta aac ctt ata	672
Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile	
210 215 220	

Figure 29

720
tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att
Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
225 230 235 240
753
tgg aat gtt gct gtt gaa gaa gaa tct tct aat gct
Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

Figure 29

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
 Met Gly Lys Leu Leu 5
 1
 ttt gct gct ggt ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
 Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30
 tca ttc tgg ctg gtt aca gct gct ggt atg tta gcg gca act gtg ttc ttt 144
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Thr Val Phe Phe
 35 40 45
 ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act 192
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60
 gta tct ggt tta att act ggt ata gcc ttt tgg cat tat ctc tat atg 240
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80
 aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat 288
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95
 att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
 100 105 110

Figure 30

att ctt gct gct tgt aca aat gtt gct gct tca tta ttt aag aag ctt	384
Ile Leu Ala Ala Cys Thr Asn Val Ala Ala Ser Leu Phe Lys Lys Leu	
115 120 125	
cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct	432
Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala	
130 135 140	
gga tta gct cct gta tgg cct gct ttc att att ggt atg gct gga tgg	480
Gly Leu Ala Pro Val Trp Pro Ala Phe Ile Ile Gly Met Ala Gly Trp	
145 150 155 160	
tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta	528
Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Lys Ala Ala Val	
165 170 175	
agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg atg	576
Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met	
180 185 190	
att att gtt ggt gga tgg gca att tat cct gct gga tat gct gct ggt	624
Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly	
195 200 205	
tac cta atg ggt ggc gaa ggt gta tac gct tca aac cta aac ctt ata	672
Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile	
210 215 220	

Figure 30

tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att	720
Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile	240
225	235
tgg aat gtt gct gtt aaa gaa tct tct aat gct	753
Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala	250
245	

Figure 30

atg ggt aaa tta tta ctg ata tta ggt agt gct att gcg ctt cca tca 48
 Met Gly Lys Leu Leu 5
 1
 ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
 Phe Ala Ala Ala Gly Asp Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30
 tca ttc tgg ctg gtt acg gct gct ggt atg tta gcg gca act gta ttc ttt 144
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Thr Val Phe Phe
 35 40 45
 ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act 192
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60
 gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tac atg 240
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80
 aga ggt gtt tgg ata gat act ggt gat aca cca gta ttt aga tat 288
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95
 att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
 100 105 110

Figure 31

att ctt gcc gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt 384
 Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
 115 120 125

cta gct ggt tca ttg gta atg tta ggt gct gga tct gca ggc gaa gct 432
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Ser Ala Gly Glu Ala
 130 135 140

gga tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg 480
 Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
 145 150 155 160

tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta 528
 Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Tyr Lys Ala Ala Val
 165 170 175

agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg atg 576
 Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met
 180 185 190

att att gtt ggt gga tgg gca att tat cct gct gga tat gct gct ggt 624
 Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
 195 200 205

tac cta atg ggt ggc gaa ggt gta tac gct tca aac tta aac ctc ata 672
 Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
 210 215 220

Figure 31

tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att 720
Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile 240
225 230 235

tgg aat gtt gct gtt aaa gaa tct tct aat gct 753
Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala 250
245

Figure 31

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
 Met Gly Lys Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15
 ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
 Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30
 tca ttc tgg ctg gtt gtt aca gct ggt atg tta gcg gca act gtg ttc ttt 144
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Thr Val Phe Phe
 35 40 45
 ttt gta gaa aga gac caa gtc agc gct agt aag tgg aaa act tca ctt act 192
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60
 gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg 240
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80
 aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat 288
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95
 att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
 100 105 110

Figure 32

att ctt gct gct tgt aca aat gtt gct gct tca tta ttt aag aag ctt 384
 Ile Leu Ala Ala Cys Thr Asn Val Ala Ala Ser Leu Phe Lys Lys Leu
 115 120 125

 cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct 432
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
 130 135 140

 gga tta gct cct gta tgg cct gct ttc att att ggt atg gct gga tgg 480
 Gly Leu Ala Pro Val Trp Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
 145 150 155 160

 tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta 528
 Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Tyr Asn Lys Ala Ala Val
 165 170 175

 agt act gca agt cct gct gct gtt aac tct gca tac aac gca atg atg gtg 576
 Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Val
 180 185 190

 att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt 624
 Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
 195 200 205

 tac cta atg ggt ggc gaa ggt gta tac gct tca aac cta aac ctt ata 672
 Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
 210 215 220

Figure 32

tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att	720
Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile	240
225	
230	
235	
tgg aat gtt gct gtt aaa gaa tct tct aat gct	753
Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala	
245	
250	

Figure 32

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
 Met Gly Lys Leu Leu 5
 1
 ttt gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
 Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30
 tca ttc tgg ctg ctg gtt aca gct gct ggt atg tta gcg gca act gtg ttc ttt 144
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Thr Val Phe Phe
 35 40 45
 ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act 192
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60
 gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg 240
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80
 aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat 288
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95
 att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
 100 105 110

Figure 33

att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt 384
 Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
 115 120 125

 cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct 432
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
 130 135 140

 gga tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg 480
 Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
 145 150 155 160

 tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta 528
 Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Lys Ala Ala Val
 165 170 175

 agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg atg 576
 Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met
 180 185 190

 att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt 624
 Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
 195 200 205

 tac cta atg ggt ggc gaa ggt gta tac gct tca aac tta aac ctt ata 672
 Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
 210 215 220

Figure 33

tat aac ctt gct gac ctt gtt aac aag att cta ttt ggt ttg atc att 720
Tyr Asn Leu Ala Asp Leu Val Asn Lys Ile Leu Phe Gly Leu Ile Ile 240
225 230 235

tgg aat gtt gct gtt aaa gaa tct tct aat gct 753
Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala 250
245 250

Figure 33

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
 Met Gly Lys Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15
 ttt gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
 Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30
 tca ttc tgg ctg gtt gct aca gct ggt atg tta gcg gca act gtg ttc ttt 144
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Thr Phe Phe
 35 40 45
 ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act 192
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60
 gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg 240
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80
 aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat 288
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95
 att gat tgg tta tta act gtt cca tta caa gtg gtt gag ttc tat cta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Val Val Glu Phe Tyr Leu
 100 105 110

Figure 34

att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt
 Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
 115 120 125 384

cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
 130 135 140 432

gga tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg
 Gly Leu Ala Pro Val Leu Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
 145 150 155 160 480

tta tac atg att tat gag cta tat atg ggt gaa ggc aag gct gct gta
 Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
 165 170 175 528

agt act gca agt cct gct gtt aac cct gca tac aac gca atg atg atg
 Ser Thr Ala Ser Pro Ala Val Asn Pro Ala Tyr Asn Ala Met Met Met
 180 185 190 576

att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt
 Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
 195 200 205 624

tac cta atg ggt ggc gaa ggt gta tac gct tca aac tta aac ctt ata
 Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Ile
 210 215 220 672

Figure 34

tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att	720
Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile	240
225	235
tgg aat gtt gct gtt gtt aaa gaa tct tct aat gct	753
Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala	250
245	

Figure 34

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
 Met Gly Lys Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15
 ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
 Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30
 tca ttc tgg ctg ctg gtt gtt aca gct gct ggt atg tta gcg gca act gta ttc ttt 144
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Thr Val Phe Phe
 35 40 45
 ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act 192
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60
 gta tct ggt tta att act ggt ata gct gct ttt tgg cat tat ctc tac atg 240
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80
 aga ggt gtt tgg ata gat act ggt gat gca cca aca gta ttt aga tat 288
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95
 att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
 100 105 110

Figure 35

att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt 384
 Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
 115 120 125

cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct 432
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
 130 135 140

ggt tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg 480
 Gly Leu Ala Pro Val Leu Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
 145 150 155 160

tta tac atg att tat gag cta cat atg ggt gaa ggt aag gct gct gta 528
 Leu Tyr Met Ile Tyr Glu Leu His Met Gly Glu Glu Lys Ala Ala Val
 165 170 175

agt act gca agt cct gct gct gtt aac tct gca tac aac gca atg atg aag 576
 Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Lys
 180 185 190

att att gtt att gga tgg gca att tat cct gct gga tat gct gct ggt 624
 Ile Ile Val Ile Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
 195 200 205

tac cta atg agt ggt gac ggt gta tac gct tca aac tta aac ctt ata 672
 Tyr Leu Met Ser Gly Asp Gly Val Tyr Ala Ser Asn Leu Ile
 210 215 220

Figure 35

tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att 720
Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile 240
225 230 235

tgg aat gtt gct gtt gaa tct tct aat gct 753
Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala 250
245

Figure 35

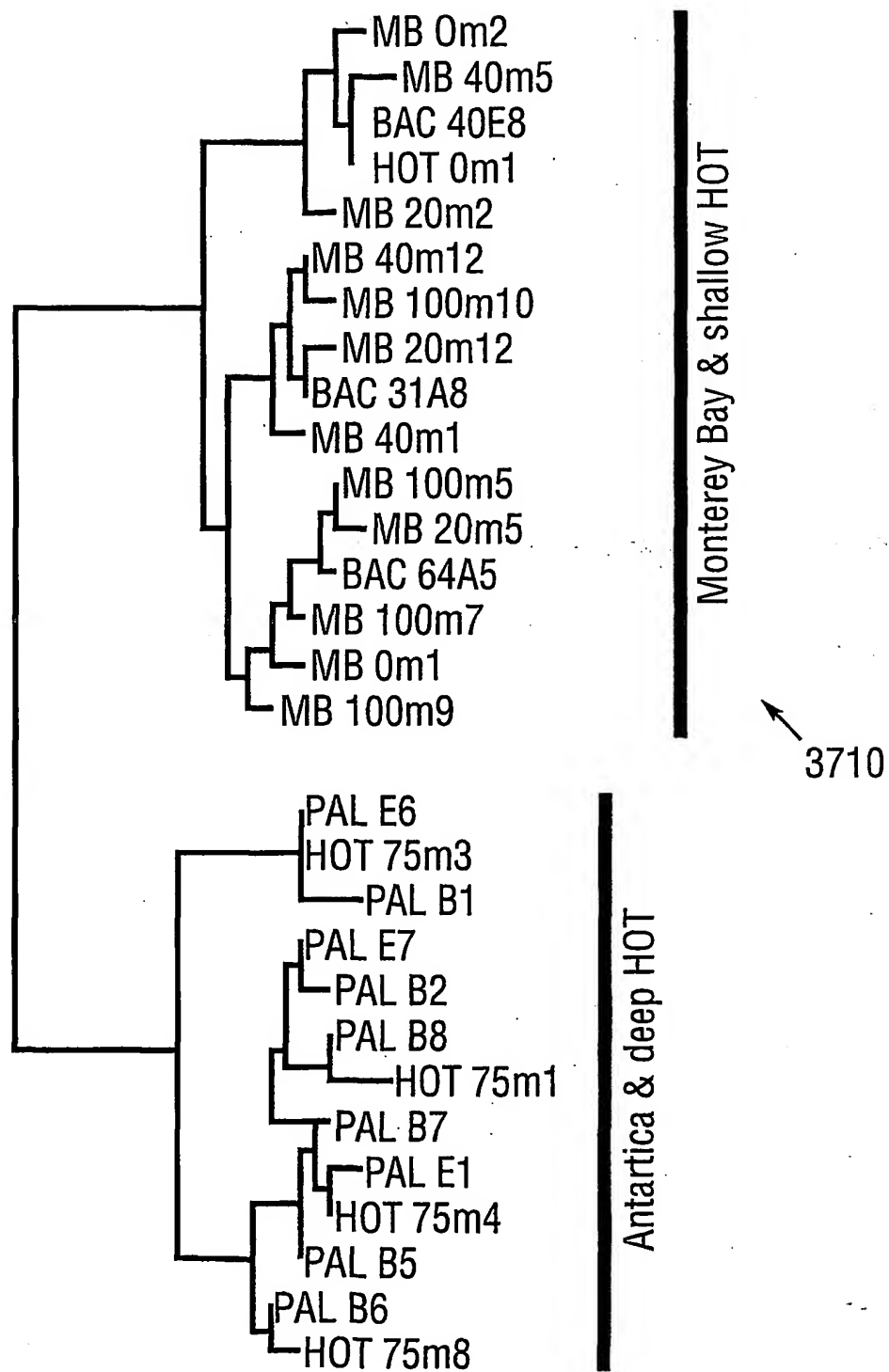
atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15
 ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
 Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30
 tca ttc tgg ctg ctg gtt aca gct gct ggt atg tta gcg gca act gtg ttc ttt 144
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Thr Val Phe Phe
 35 40 45
 ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act 192
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60
 gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg 240
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80
 aga ggt gtt tgg ata gat act ggt gat acc cca aca gta ttc aga tat 288
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95
 att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
 100 105 110

Figure 36

Figure 36

tat aac ctt gct gac ttt gtt aac aag att cta	ttt ggt ttg atc att	720
Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu	Phe Ile Ile	240
225	235	
230		
ttg aat gtt gct gtt gtt aaa gaa tct tct aat gct		753
Trp Asn Val Ala Val Lys Glu Ser Ser-Asn Ala		
245	250	

Figure 36

**Fig. 37**

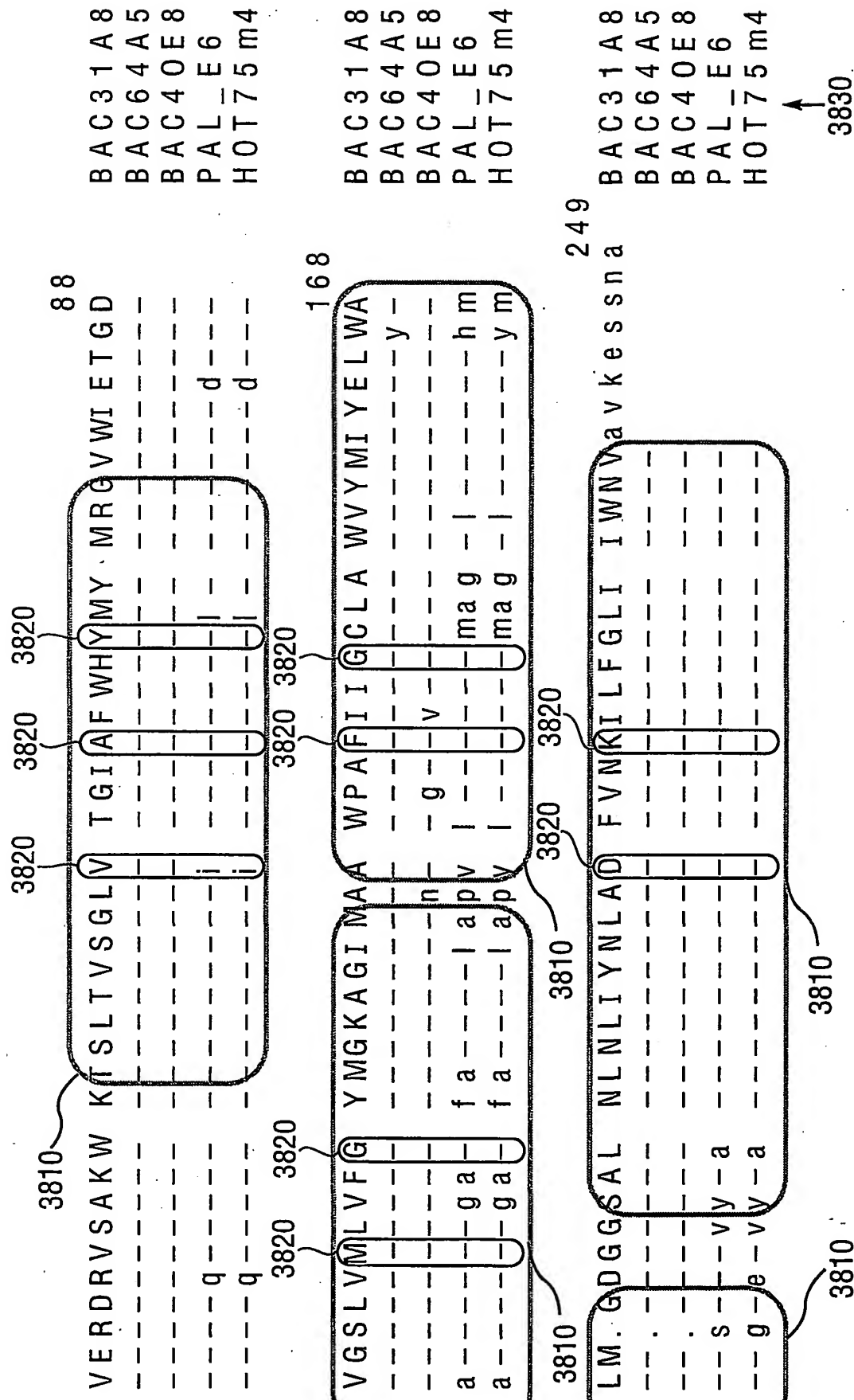


Fig. 38B

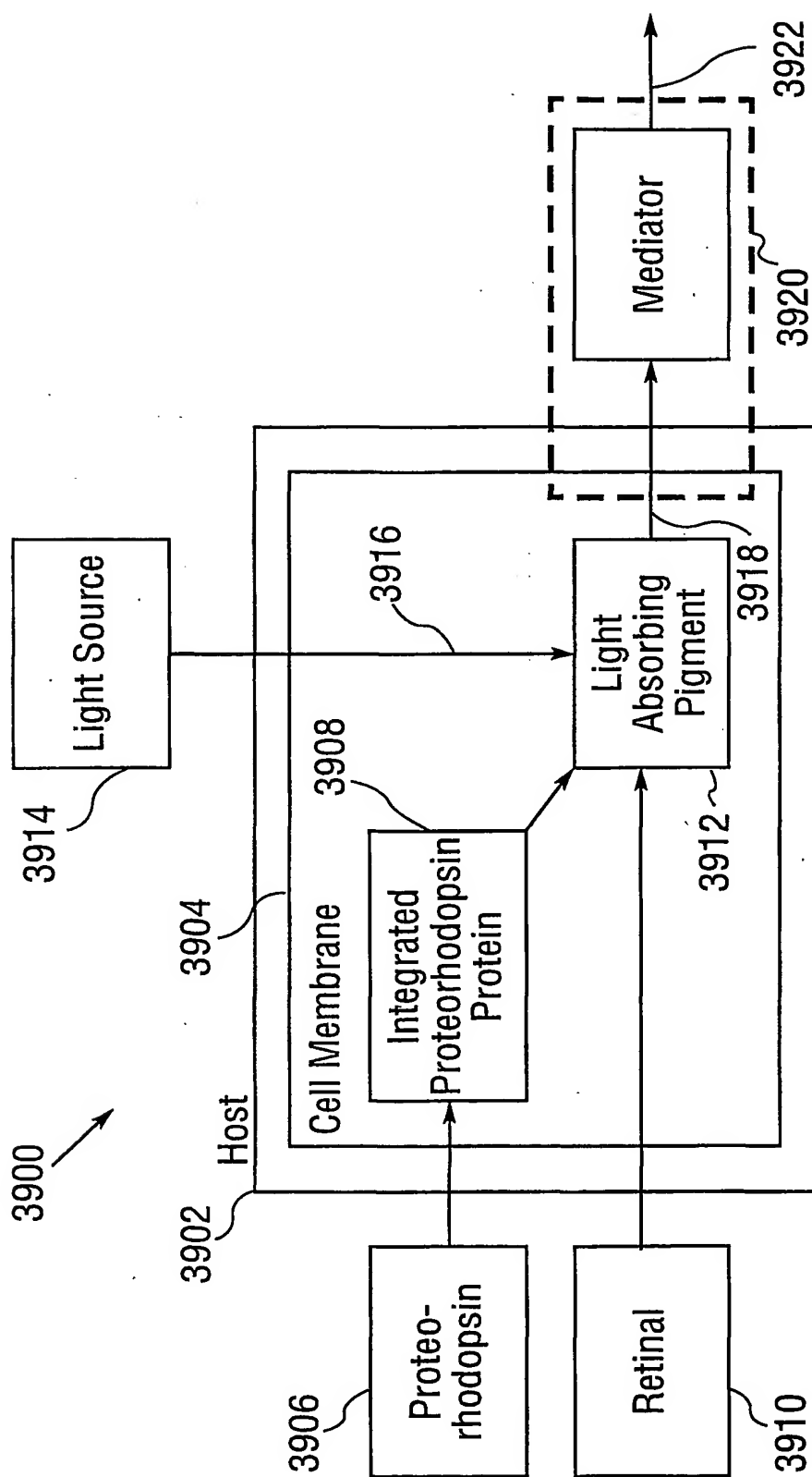
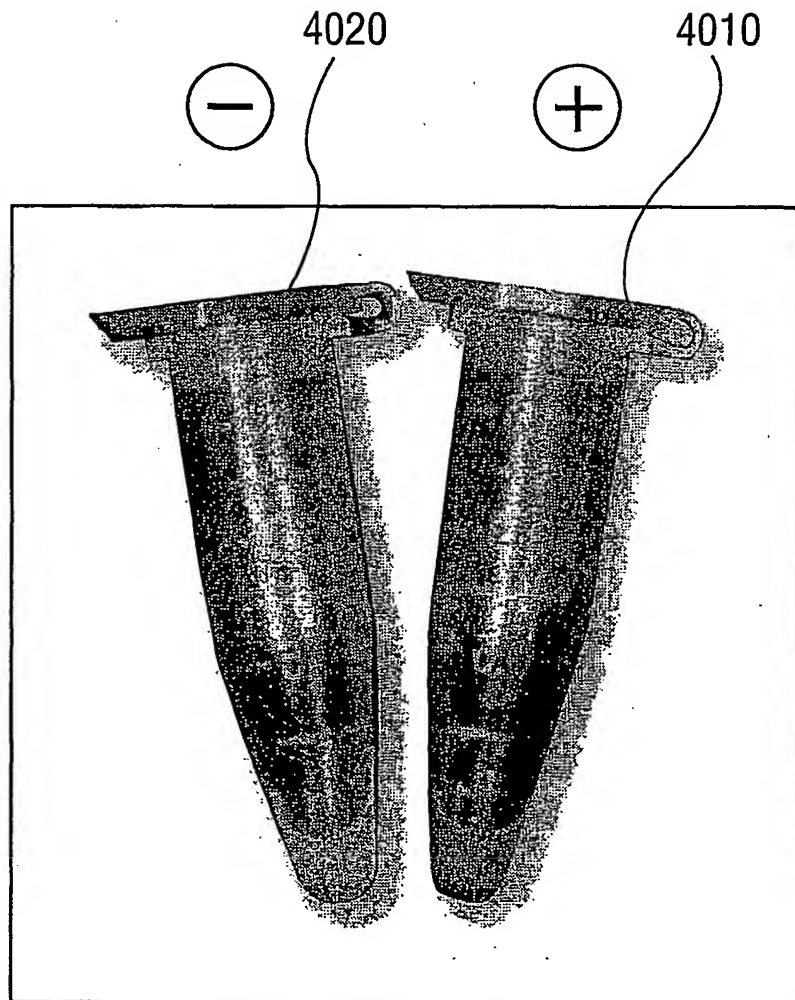


Fig. 39



BEST AVAILABLE COPY

Fig. 40

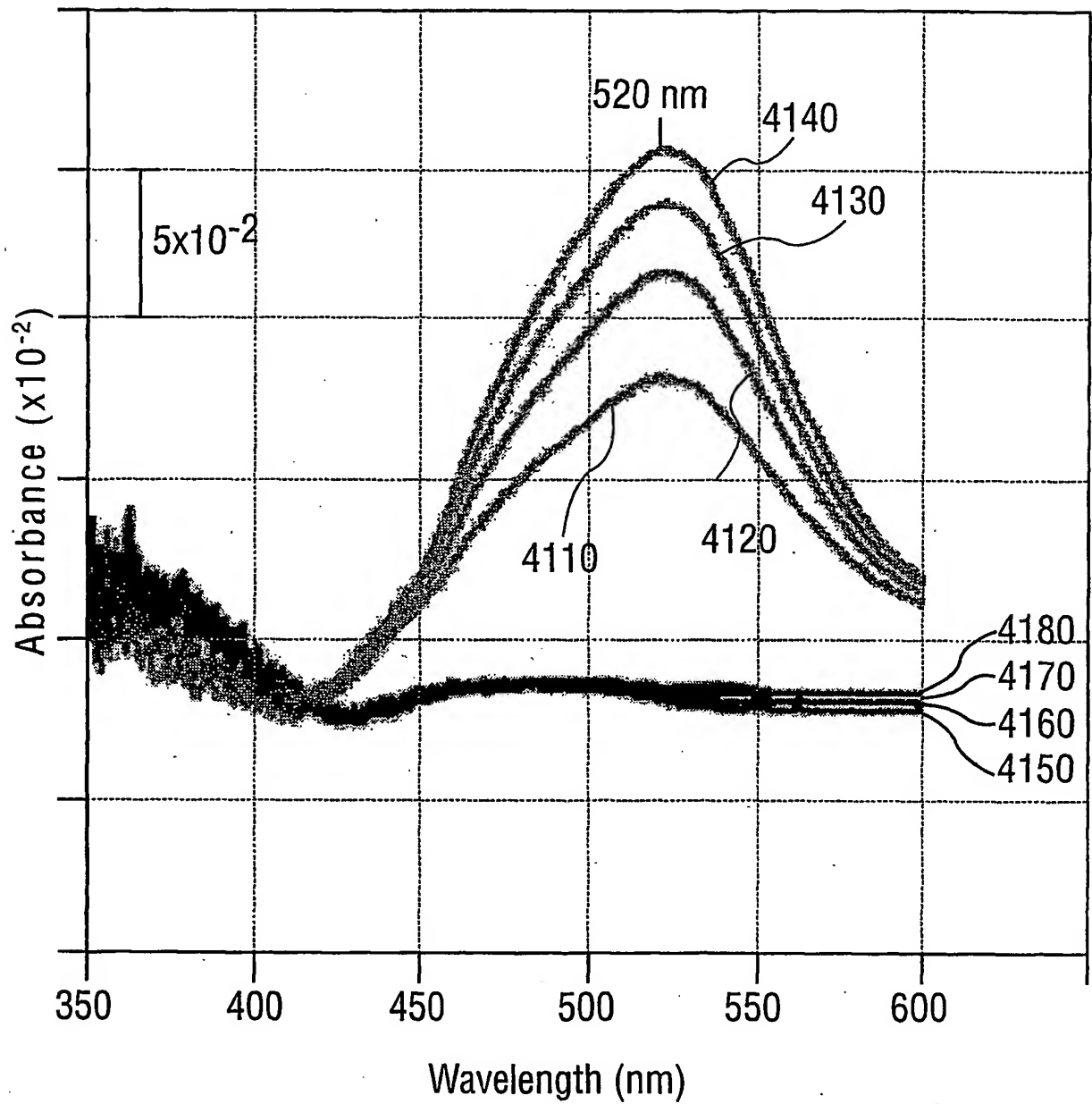


Fig. 41

BEST AVAILABLE COPY

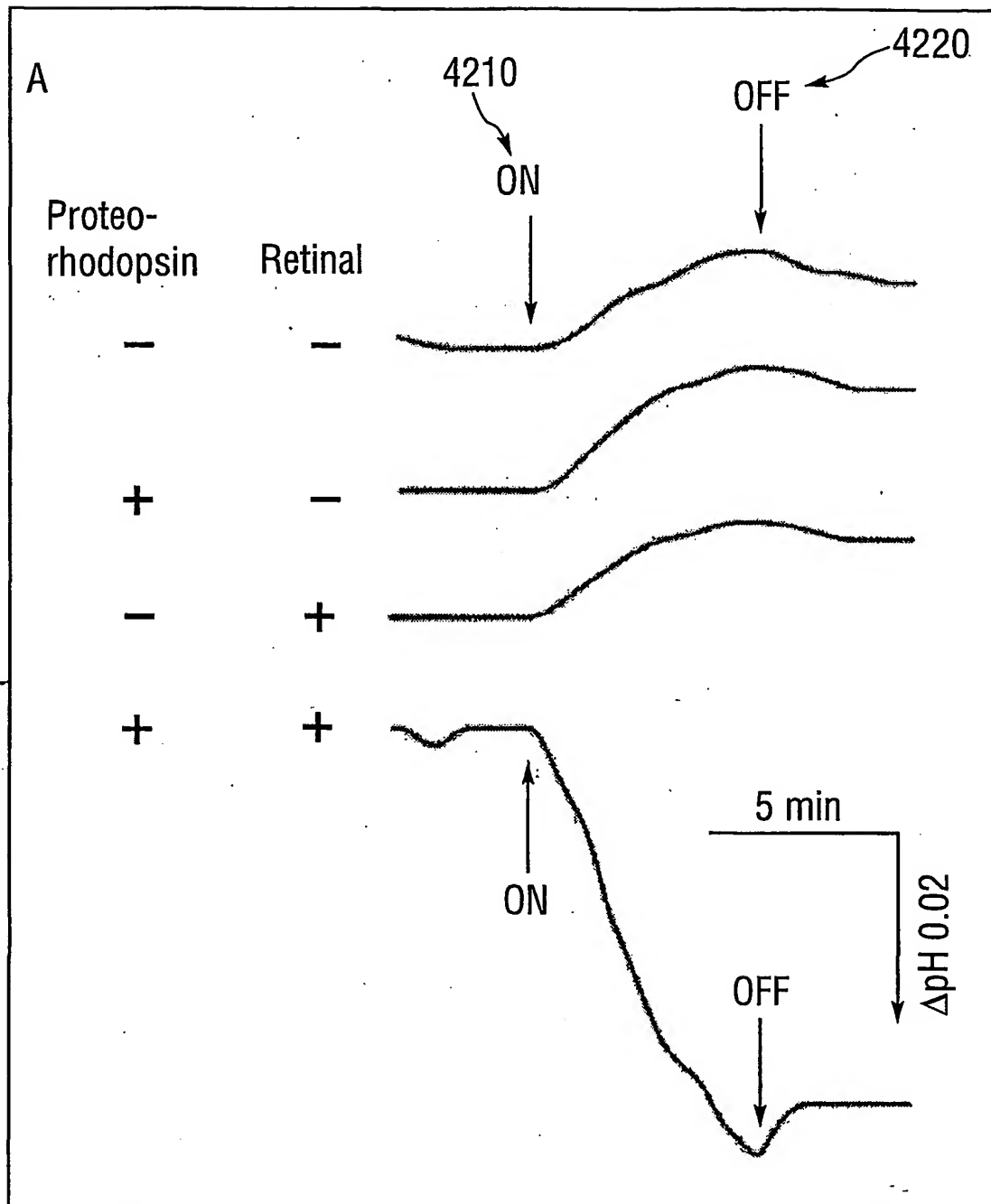


Fig. 42

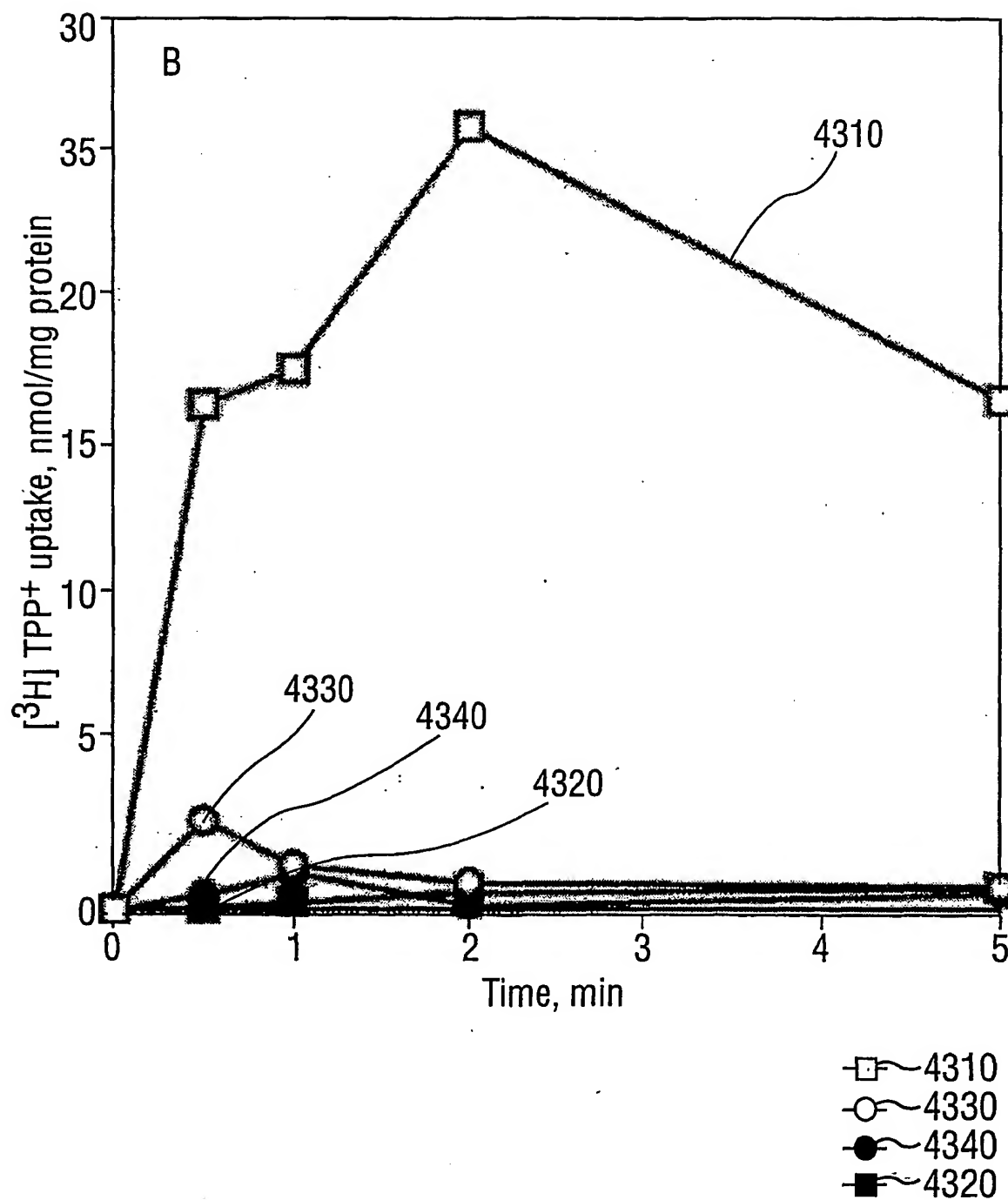
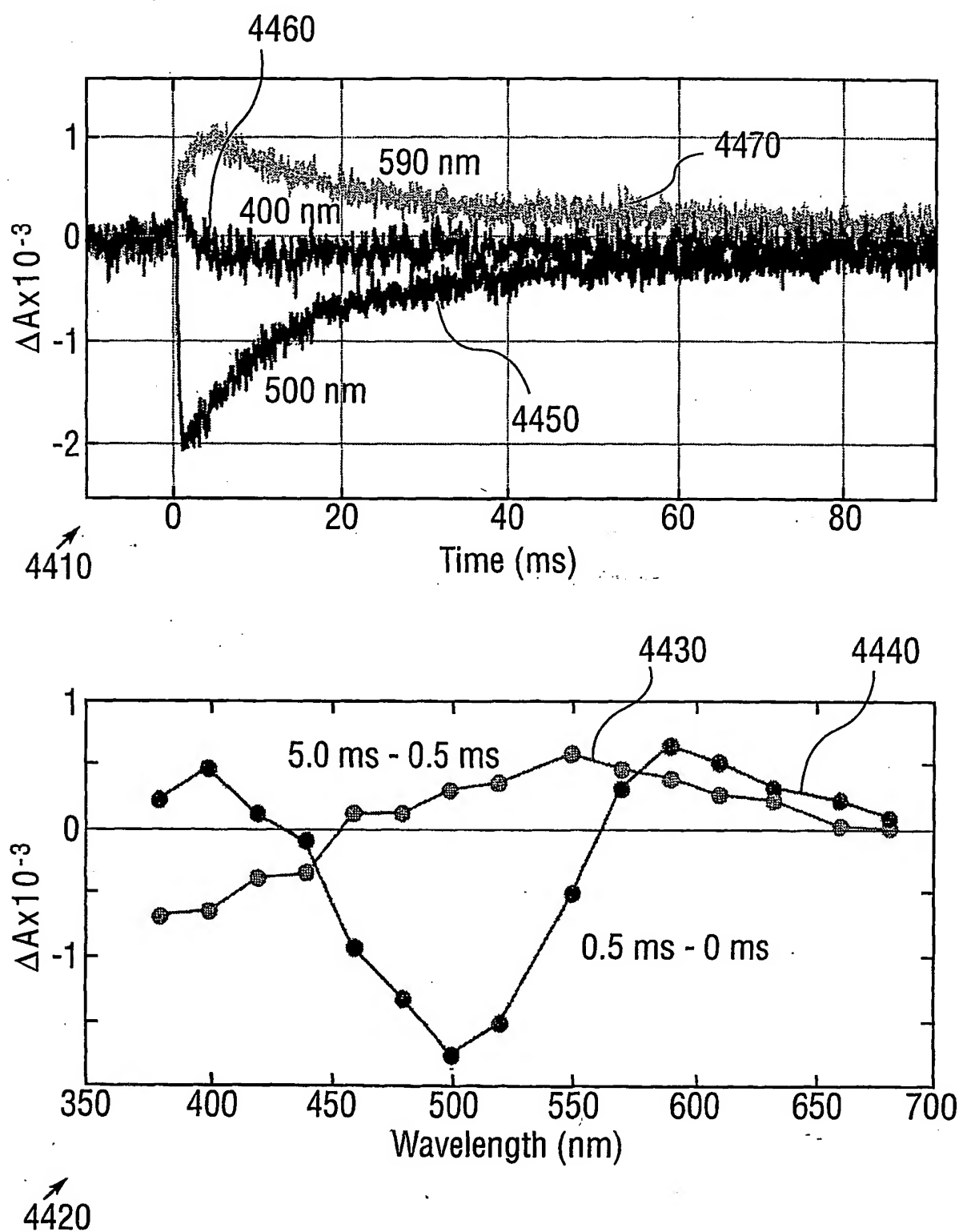
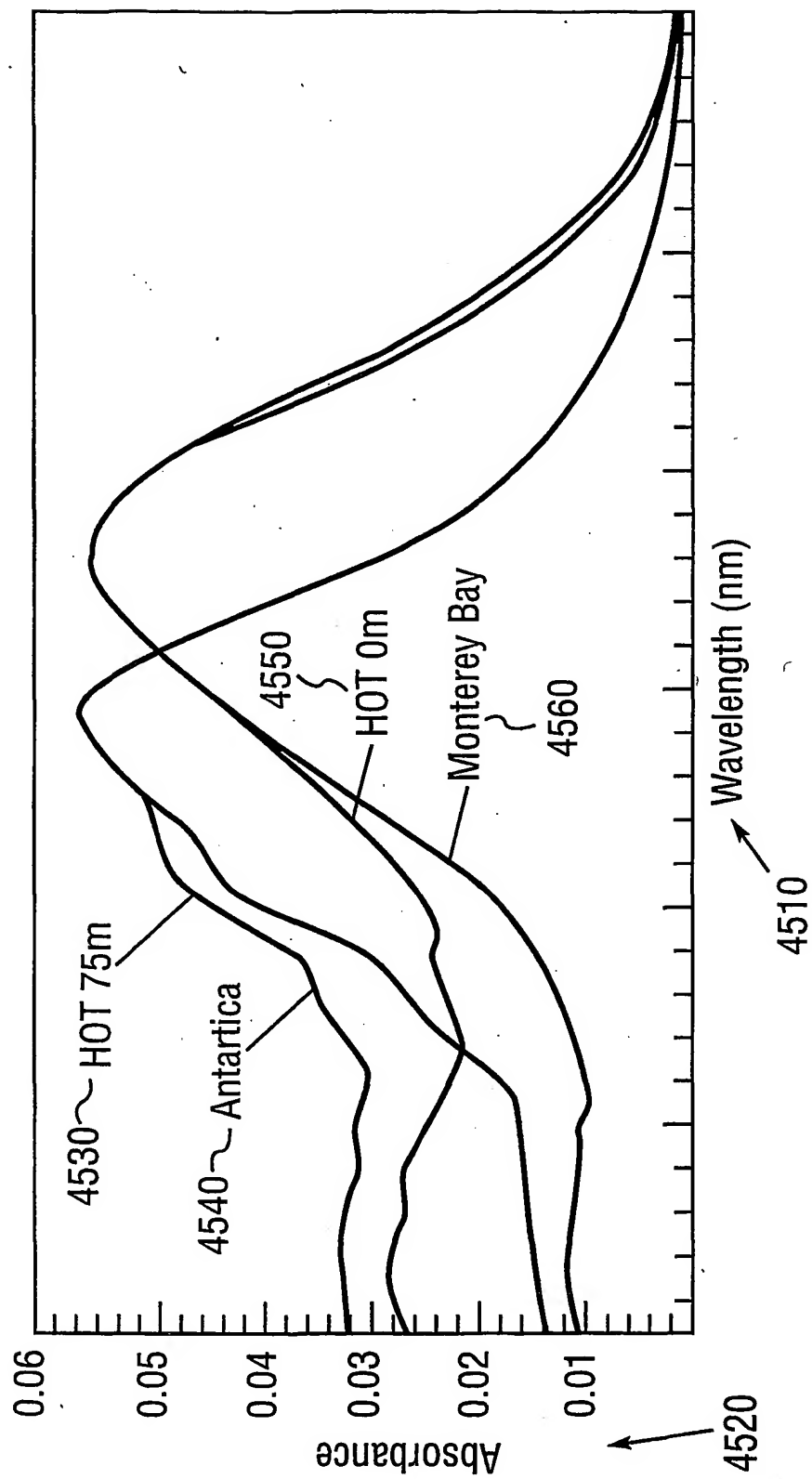


Fig. 43 BEST AVAILABLE COPY

**Fig. 44** BEST AVAILABLE COPY

**Fig. 45**

SEQUENCE LISTING

<110> MBARI
DeLong, Edward
Beja, Oded

<120> Light-driven energy generation using proteorhodopsin

<130> MBA-101

<150> 60/201,602
<151> 2000-05-03

<160> 65

<170> PatentIn version 3.0

<210> 1
<211> 105184
<212> DNA
<213> Naturally occurring gamma proteobacterium

<220>
<221> CDS (complement)
<222> (50866)..(51615)
<223> light-driven proton pump; has the properties of a light-driven proton pump when expressed with retinal in Escherichia coli

<300>
<301> Beja, O., Aravind, L., Koonin, E.V., Suzuki, M.T.,
Hadd, A., Nguyen, L.P., Jovanovich, S.B., Gates, C.M., Feldman, R.A.,
DeLong, E.F
<302> Bacterial rhodopsin: evidence for a new type of phototrophy in the sea
<303> Science
<304> 289
<305> 5486
<306> 1902-1906
<307> 2000-09-15
<308> AF279106
<309> 2000-06-15
<313> (50866)..(51615)

<400> 1
ttgttatatc agtaatggct attgctccaa taacttaata ctaatatata attagtttat 60
gaataaattt tatatatattg gggtattgtt ttttacacta aatgcatttt cttgctcaga 120
tcttctagat acagacatga gagttcttga ttccgctgag tcaagaaacc ttgctgagtt 180
tgaaggaaaa gctttactag ttgtgaatgt tgcaagtaga tgtgggttaca cttatcaata 240

tgctggcctt caaaagtat atgaaagtta taaagatgaa gattttctag taattgggat 300
cccatctaga gattttcttc aagaatactc tgatgaaagc gatgttgagc aattttgttc 360
tacagaatac ggtgttgaat ttcttatgtt ctcaactgct aaagtcaaag gaaaaaaagc 420
acacccattt tataaaaaac ttattgcaga atcagggtttt actccctcat ggaactttaa 480
taaatactta atctcaaaag agggcaaggt tgtatccaca tatggatcaa aggtaaagcc 540
tgattcaaaa gagcttatat cagctataga aggcttgctg taaaattatt acttagaaac 600
taatacagtt ttaggcttgt ttgctgcaa tattccatta tctacaactc caggaatatt 660
attaatcaaa gcttccattt cagtgggggt tgaaatatcc atattagaga tatctaaaat 720
gtgattacct tgggtctgta taaatccagt tctatatgtg ggtattccac cgatcgagat 780
tatttttctt gcaacaaggc tctactttc aggtatcacc tctataggca gtggaaaagc 840
tcccaaaaga ttaaccatct ttgactgac aactatacat ataaactcgt tagaggcaga 900
agcaactatc ttttctctag tatgtgcgc accaccacct ttaataagac aattttcagg 960
agacacctca tctgcaccat ctatgtaata agctatatca actacatcat taaggctaaa 1020
gacctctatc ccattttcat ttaataattt tgatgaagca tctgaactag aaacagctcc 1080
agcaaatttg tgcctatgct ccttttagttc ttctataaaa aaattaactg ttgagccggt 1140
tccaatacct aaaatcatct caggatgaag attatttttg atatattcta tagcttggtt 1200
agcaacattt atctttgagc cactcataga gttataatac aagaaaatat aggtagttaa 1260
ttattttgag actaaaaatt aaaaaaacag gttcttttaa gaattcccag aagtacctaa 1320
agcttattcc taatgcatct gtttatgacg tagcattaaa atcacctata acatttgctc 1380
taaataattt tcaaaagctg gggaataaag ttttctctaaa aagagaggat ctgcaaccta 1440
tatttttctt taaaaacaga ggagcgtata acaagattgt aaatttatcc gatgccgaaa 1500
agaagagggg ggttattgct gcatcagcag gaaatcatgc tcaaggggta gccagtgcac 1560
gtaagaaatt aaaaattaat tgcttgatag ttatgccaat aacaactcca gaaataaaaa 1620
taaaagatgt aaaaagattt ggagccaaaa tactccaaca tggggacaac gtagatgcag 1680
cattaaaaga ggcaactgtt attgcaaaga aaaaaaatt gtcttttggt catccttttg 1740
acgaccctct aacaattgct ggccaaggga ctataggaca agaaattctt gaagataaaa 1800

ataatTTTga TgtTgtcttTt gTtcCGgtTg gaggaggagg tattctagct ggtgtatctg 1860
cctggatagc acagaataat aagaaaataa aaattgTtTg TgtTgaggTt gaggattccg 1920
ctTgtctTgc Tgaggccgta aaagctaata aaagagTtat tttaaaagaa gtgggcctct 1980
ttgctgatgg ggtggcagta tcaagggtTg gaaaaataa ttttgatgtt attaaagagt 2040
gcgtagatga agtcattaca gttagcgtTg atgaggTctg caccgctgta aaagatatct 2100
ttgaagatac aagggtTcta tcagaacctg ctggggcatt agcacttgca gggTtaaaag 2160
cctacgcaag gaaagtTaaa aataaaaaac ttattgctat aagtTctggc gctaattgtaa 2220
atttccaaag acttaattTt attgTtgagc gatcagagat TggTgaaaat agagaaaaaa 2280
tattaagtat caaaatccca gagatacctg gaagTttTct taagctTtca aggatgtTtTg 2340
gcagctctca agttacagag tttaaactaca ggaaatctag cttaagcgat gcatatgtTt 2400
tagTtggtgt tagaactaaa actgaaaaat catttgaaat cttaaagtcc aaattaaaaa 2460
aagcaggctt cacctTtagc gactTtactc gaaatgaaat atccaatgat catctgaggc 2520
atatggTtTg Tggcagaaat agtgactcag gctctcataa caatgaaaga atatttaggg 2580
gagagTttcc Tgagaagccg ggcgcgctgt taaattTtct agagaaattt ggaaataaat 2640
ggmatattTc ctTattTcat tacaggaacc taggtTcagc tttTggaaag atattaattg 2700
gcatcgagag taaggataaa gacaagctaa taaatcattt aaataagtca ggnactattt 2760
ttacagaaga aacctctaac aaggcataca aagattTttt aaaatgaaag gTtaatactt 2820
taatctaaat ttaattgaaa aaagctcatc gctagggtTt tcccacggct cttTgaacaa 2880
ctcggattga gatctatcat cctcctcgtc gTaaattctc ccacctTtag aatagaccaa 2940
aaatagatat gacaaaggag cgagctcata tttatatcta attTggaacg aagctacgcc 3000
agtattaaat tcattaacta tattattTcc tttataaagg tatccattag catctgaaaa 3060
aatactaata ggattTttTg cttTcaaagc aacaaattga ctcttaagtc Tgatctcatg 3120
tttattattt ttaaaccaat ttagatcaaa agataaggta tctTgtctTg aatcatatga 3180
ggcaagatta ttattatcct gccatatcag ccattcattt tcctTtctta ttctatattg 3240
cgcattgatt cttaagTtat cattTggaaa tattgaacct gctatctTgt aaaattTtct 3300
accatacca ttcgaatccc accgattatc tttctctccc ttaaagaagc taactctcca 3360

gtcatatgtc cagaatgaat agttctttgc ctcaaagtct gctgtaatac ctattcttct 3420
ttttgatttg ataaaggggt aggettccatt ttttcttgtg atagttgtat ttttcccaga 3480
agatctaaag ttaaaatcta attgaaattt agagttgtcc ttaaaactaa aagaattttt 3540
ttgatcgatg cctattggat ttgaattacc agctgtgtca gcatcataat ttagatcaat 3600
tccataatct atttgtttta atacgagct attatcaaat tcattttattt ttcgatttcc 3660
tccaatacca gcatgaatcc agtctcttct ttgcagataa ccaaagtcatt ttaactcgaa 3720
gtcgtcttca aaataaagaa ggcttccact tatgttagat agttttatttg gaagatatgt 3780
aaactgagtc ctatacccaa gcccatTTTT accatctttt tctgaggcta acagatctga 3840
atatgtaatt aatttttttg aacgaatatt gatgtaatca ataacattga ctgttgatga 3900
ttcgctgtc atttcattct caacattcgt taccatgaag ccaagcggtt tatttccaag 3960
ctttgttcca gatcgaagag cataatagtc tcttccaact gaaaaggctt catcagcctc 4020
acttgctaca aatactccaa attcattatt attacttttt tgagtaagtc ttaatgcaaa 4080
atcaatatca gaatagtttt ttttagctgc ctgcgaacct tcttcattac tttcttctga 4140
gcaattatag ctgggggagc ctccaatcct ccgcgtattt ataaccgagt atctatcata 4200
attactaata tcaaatagtg attgggtttt attgaaaaat gctctttttt ctgagtaaaa 4260
agtttcttga gcagaaaagt taataaccac atcatcactc tcagcttgtc cgaaatctgg 4320
attaatagct aaattttatt gacgacctt tccagtgtca taaaagattt cagccccaat 4380
atctgacctt tcttggtttg taactgaatt tttatttgaa gatatatatg gaaaaaagt 4440
aagctttgat tttgtatagt tttgtatttc taagctatct aactcttgaa agtagtcatt 4500
tctactagcc attgttccgg cactgctaac ccatgactca ttttcggcac tataacgtaa 4560
tgcggtgtaa ttaatttttc ttatatcacc atcaggtgtt ttcattaacg ttacatccca 4620
aggaataaaa aactcagaga cccaataccc atcaaatttt tgtgtttttg caatccaatc 4680
tccatcccag tctgtcttaa agtctcctgc ttgcgttttt atggcatcga aaagcgagtt 4740
cccaagattt atagcaagaa tgaaagcttt gttaccatca ccatcaaagt ctatatttat 4800
agagttttta tcgctaagtg aattttatttg atctctaagc gtcttctctgg agaacataga 4860
atcattactt tgaaaattct taaatccaac atatatacca tccttatttg agaaaattaa 4920

agccgttgta agaagttcat ttttcttaag agtaaaagga gatgtttcat aaaaatctgt 4980
aatttcaaat gcattattcc actcaggctc atcaagagag ccatcaataa caattgagtt 5040
tgaccaaadc agcacagagg taagtaaaag tgataatgag gctaaagtta ttttcataga 5100
tagatttaat tcagagtatt ttaatctatg aagagtagga aatcatccga tcattctcag 5160
aaaaaacata atcatattta aggtcatgtt tttctccaaa atgcatatta caaatctggt 5220
attgataaca cagtccaacc aataaaggctc ttgatacctc ctcgtttata gagccaataa 5280
ttctatcgaa atatccagag ccatagccaa gcctgtatcc atttaaataca actcctgtca 5340
taggaataaa cattaaatca atttcattta tgttgacata atcttcactt ttaacctctt 5400
tgatcccaaa ttgatttata aagaagttag gctgctcatc cagcaaatta aaagccatca 5460
tttctgctatc aattactttt ggaatataaa tattcttttt aagtttagta aaagcttgaa 5520
ttaacaaatt tgtattgact tcatttcgaa aaggaaaata taaggcaatg ttttgcattt 5580
catgagtatt gattttttct aatacatttt cttggattaa aaaacttata ttgctcttag 5640
ataaatcaga aatagactgc ccctgctcaa agagtgattt tcgtatttta tttttcacca 5700
taaattgtgc cgaggctaac aaaccaaadc accgccatgt gctatacctg aaccgatggg 5760
tcaggtggtg aatttcataa caaatcaggc ttccctgtaa aggtactgcg caacaatgct 5820
agaggaatat cacctattta attgtatcgg ttcaaattta ttaacacatt agcgtatgaa 5880
ccagaatgca gttaattata aaatatatag gtattaagta aaagttaatt tttagagagc 5940
agactctatt ttttgtatta gcttttcgat atctttatta tttaaagcag agtcagtatt 6000
ctcggttggc ttggataaaa gcttactagc catagttaga ccagctatca ccaaggcatt 6060
attcttatca ctgatgccat caagctcatc atttaataat tttgcagctc taatgagttg 6120
gtctttctct tctggcggac aggctaacgt cagatctcgg ccaaaaattc ttaaagatag 6180
cgtttccatt tctgacatta ttttttgagt ttagttattt ctttatttag tatgagctta 6240
tcttttttcc agtctctttc cctttgcaca taagaatcta ttattccttt ttgctcatca 6300
aatcttttta tcaaaagatc taccttatcc tctagttcaa ataataatga tttttttctt 6360
tctgacattg tataagttta attgtactag aatgaattgg aaagtttctt taataaaaga 6420
gtaaaataag gcatggaaaa aataattttt aaaaacagaa gagattcact tataaaacac 6480

ctacctaaga attcagcctt aattgtacct ggtgcagatt tgcaatatag aaatgctgac 6540
tcattcttata atttaagaca ggaaagtagc ttctactatc tgtctggctt ttgtgagccc 6600
tcttctctaa tgggttttagt taataatgga aaaagcattg attcaataat ttttgttcct 6660
gaaaaagata aacttaaaga aatctgggat ggttatcgag ctggccctga gggcgcaata 6720
aatgattttc tttttgatca agcttttgaa aataataaat cagatgcttt aatgcctgaa 6780
atccttcaag ggctagaaaa agttttttat tcaataggga agaaaaatgg ctttgatcag 6840
aaagtaattg actggacatg cgcagcaaat tctaaagata ggcacagcaa atcaattgat 6900
attattgatg gctcttcgat ggtaggaaat ttaaggctta tcaaagataa gcatgaaatt 6960
gatattatga agagagcttg tgaaatttca gctgaatcat atattgaggt catgaaatct 7020
ataaagcctg gggacaatga gcaggaaata gaggcgctat ttttatatga attcgccaaa 7080
aggggaggaa ggtttccagc ttatacacct atagttgctg gaggtgaggg tgcttggtga 7140
ttgcattata ttgaaaatga taaagagtta gcttcacagc atttaatttt ggtagacgca 7200
ggatgtgaat acaaaatgta tgcattctgat atcacaagaa ccttcccagt aagtggaaaa 7260
ttttcagatg aacagctaca aatttataat attgtccaca aagccaatct tgctgcaatc 7320
gatgctgtaa aaactggtaa tagcataatg gagcccaaaa tggtttcaga aaaagtaatt 7380
actgaaggtc ttgtagagtt gggatttcta tctggcgatg ttaatcagct tcataaaaat 7440
ggtgcattca aggactttta tatgcataag gtgggacatt ggcttggact tgatgttcat 7500
gatgttggtg actacatgga gggagatgag tttatgaagt ttaagccagg gatgataacc 7560
acaatcgagc caggcatcta tatcagtagc gcaatggatg tagatgacaa atggaaaggc 7620
atcggcataa gaatagagga cgacatcctt gtaacagatt caggcaatat taatctaaca 7680
gagaagggtc catctaatec tcaagaaata gaatcattga tggcttagac tatggagggt 7740
ccaattgtta tttctggcgg agggataata ggtaattaca tttctcttag gcttgaaaaa 7800
aataatatca aaaccgttat tgtcgaaaaa gctagtagtt tcaaagccct agataagggt 7860
ataagaacag tcaactctcaa tgagcattct atgcaaatgc taaaaaatat tggattttgc 7920
ccatcaattg ctcaaatcaa cagcatcgac gtattagatg gtgagggtac aggcaaaatt 7980
caatttctag caaaggacgt aggcagcgaa aacctttcat atgtaaccta tttcaatgaa 8040

ttacaaaaac taattttctga tccatgtaaa gaaagaacct tatttaataa tgagattgat 8100
tcagttcaga atcttaatac agaattctgat ccagagatca tgcttaaaga tggcatgacc 8160
ataaaaaacga atctaattgc tggatgtgat ggaagaaatt caaatattgc aaaaattgct 8220
tcacttacaa gcagcttcga tgactactta caaacagctt taacttttgt cgttgatatt 8280
gataatgatt cacatggcaa agctcaccaa gttttttctg aaaaaggaat atttgcactt 8340
atgccactcc cagaaggcaa ggggtgagatg aataaatgca cagtggtttg gtcaataaaa 8400
aatcaagttt tgggagatga gcctgtatct gagtttgtaa aaaataacat ttcttttttt 8460
gaatcaaagc ttaatgttag tctcagggtt aagtcagaaa ttttaagttt taaattatcg 8520
aaccatcatt ttgaaaacta tattagcgga cctattgttc ttcttggtga tgctgctcac 8580
tcaattcacc ccttagcagg tcaaggattt aatctaggat ttgcagatgc agatactttt 8640
tgtgaagagg taattagtct ttataaaaaa gggattgcct ttaatgagaa atcagtttta 8700
aaaagatatg agattagaag aaaaagtatg aactttttaa tgttgaagtc tatggacttt 8760
tttgtggatt tatttggttc agaaaattta tatctaaggc tgatcagaaa tttaggattt 8820
tcttcgttaa ataaatcaaa gtttggttaa gcattcttta taagacatgc ctctggaatg 8880
aataagtttt aaatttgtat taaacttttt gacctttagc tctaagttct ttaagaactt 8940
cactaatgcc ttttttatca atgattctca taccttttgc agatacttta agatttacga 9000
acctgttctc agattcaacc caaaatttgt gtgtgtgaag attaggaaaa aactttcttt 9060
tagtcctatt ttttagcgtga gaaacattgt ttcttgactg tggatatcta cctgttactt 9120
gacatatttt actcattgaa acgcgatttt atagaacact gaggaactta gcaatactat 9180
tgtgaaacaa atttatttat tacggcatgc acaatctgac tgggagagct ctaatcagaa 9240
agattttgat agaccattag caagaaaagg cattgaagaa gcaaataaaa tatcatgtta 9300
ctgcaaactc cattcaattt tagtagataa aatattctgt agcactgcag aaagaactaa 9360
gcagactttt gatatatgca gtgatgggct taattatcca atagctgaag cagtctatac 9420
tgatgagctt tacttttctg gccctgggtga gatcggttaag cttatccaaa gtttaagtga 9480
attcatttcc tctgttttaa taataggcca caatccatca atgcaaattg acatagatgc 9540
tatttcagaa aatcctcata ttacgtattc aacatgcggg ctggcagaaa ttctcgttga 9600

aagttcatgg aaagacttat ctttaaaaaa atgtaagtta aaatctttta ttcaaccagg 9660
agagctttta aaattgaaaa acataaaaat taagatcatt aaccactaa tgggatccaa 9720
gataccctta cctcaatatg aaacaaaggg ctgggcagga ttggatttaa gggcatgcct 9780
agatagtaat ctgagccttc aagcaggaac atctcagttg atacctattg gttttgcaat 9840
gtacttagaa gatcccggtc ttgcagcaat gggtatacct agatcagggt taggttctaa 9900
gcatggaatc gttcttggtta atctggttgg gttgattgat tcagactatc aaggagagct 9960
aatggttcct gcctggaata gatcagatac agattttgag attaatcctg gagacaggat 10020
tgcacaaatg attatagttc cagtgattca agcagatttt gaaattgtag acgagttcaa 10080
tgagactcag aggggagaaa agggttttgg aagttcaggt ataaattgat aaatttactt 10140
tttcttgcca aatctttctt caaatttctg aactcttccg ccagtatcaa taattttttg 10200
cttaccggta taaaaaggat gagaagcaga ggatatatca agagggtagt atgggtatgt 10260
ttttccatct tccattctt tcgtttgagt cgtatctaata gttgaacgaa tgagaaagaa 10320
cttatcagca ctagegtcat ggaataaaac ttcacgggtat tcaggatgta tatctttttt 10380
cataataaaa tcaaaatttg gatgagaact ataacaaaaa acaacttctt ttcaatcaaa 10440
acatgaaaat tttttactat gacatagctg tttcattgcc tctaaggcaa tgctttactt 10500
atagttctga gcttaaaatt acaaaaggaa cccgcgtgtc agttcctttt ggtaaaagga 10560
agattgtagg agtggttata aaaaatatc aaaagccaga tttcctaaaa aaagcgggag 10620
ctataaaaaa aattatcgct gtacttgatg aatatccttt gttcgacaag ccaatttttg 10680
attctatatt gtggtcttct gattactatc atcatcctat tggtgagggt tttaatacct 10740
ttatacccac cgaattaaga aaaattaata ataaaaaat tgaagcttta agagaatttt 10800
ctgaatatc agtaaagag gatgataaga aattcgattt aaccaaggat caagaaaaag 10860
cagtcaaggc gcttttctaa tctaaaggat tttcaccac tttattatat ggagttacag 10920
ggctctggaaa aacagaagtt tacttaagag ttgccgaaac ttttattaaa aataataagt 10980
cagtattagt tttagttcca gaaataaatt taacccccca attgctctcg cgatttgaga 11040
atagatttaa tggtgaaatt ggcataatc attctaagca aacagcagct aagagattaa 11100
agacttggct aaaagctaaa tttggttcta taaaaataat agtaggaact cgatcttcgg 11160

ctttagtgcc tttagataac attgggttaa taattatcga tgaagagcat gaccaatcat 11220
ttaggcagtc agaagggttt aaattctctg ctagagactt aagtataaaa agggcacagc 11280
ttgcagatat tccaattatt ttgggatcag caacccttc gctgcaaact ttaaaacttg 11340
taaaagaaaa taaatttata agagttgata ttcctaatac agttgatgga aacaagcctc 11400
ctaaattaat agccttagat atcaataaca gccctttaat aggcggagtt gctaaagaga 11460
caattgaagc aatgcaatca accatagaca gaggagaaca ggttctagtt tttattaata 11520
gacgaggatt cgtccactc tatcaatgta gtagttgtgg gtgggtagca gattgtaaat 11580
cttgtgatac aaatttagtc ttccaccagg caagaaatag attaatttgt cataggtgtg 11640
aatctgccta ctctgttaat ttgtcttgtc cggcatgcaa gtctaatac tttaatatgt 11700
atggagctgg aacagagaga gttgaagaag ttcttaaaag cagctttgta aagactccaa 11760
taattagagt tgatcatgac tcaacaaaaa aagtgggagc tatggaggct atagttaaaa 11820
aaattcattc ctcagacgca gcaattttag ttggaactca aatgcttgca aaaggacatg 11880
attttcctaa agtcacctta agcgttattt taaatgctga taatggcctt ataagcccag 11940
aaattaatgc attagagaaa atatctcaat tgcttattca ggtctctgga agagcaggaa 12000
gaaataataa tcttgcaaaa gttattattc aaacaagata tcctgatgat ataaatctta 12060
ataaaattaa gacaggagat tatatgaaat ttgcttctca atgcctaagt accaatgagc 12120
aatgaactt acctccattt actactttat gtctgcttag gtgctcatca ccaactcaaa 12180
agagtaatgt agatttccta gagaaagctg ttttaatttt atccaatagg actgatataa 12240
atgttattgg tcccttgctt tcattagttt cgaagtcgaa aggaaattat aggcaccaag 12300
tctatatcca tgcaccaaag aagacttttt taaataaggt attaaagttt ttgacaacag 12360
agtttgaaaa atggccggaa tctaataagg ttaagtggtc tttcgacatt gatccaatag 12420
acttaagcta aatattaatc ttaattaatt gtctgggta tattggttta ttgtttagtt 12480
tattctctgt attaatttct tctacagtca ccccaaactc tatcgctatt tctgataaga 12540
catccccctt ttgtattttg taagtcacaa agcctggatc aatactcata aaggattttg 12600
ctttagggtt gtcttaaaaa tagttatgta ttcttaggaa aacagatctt gcaatcattc 12660
ttctccctgg cttgcccctt aatctttgtg cgtcttcagg gttgggtata aaccctgact 12720

caaccaatac agaaggtata tcaatagact taagaactct gaaatcagcg tactcaacat 12780
tctttttatg aatttttgtg aaagggctct ttttaagttg atccaatata ttagttccta 12840
aaattttact ttcagaaatt tttttcttat atattttctgg ataggtttgt cttgccgcat 12900
cctcatcaaa atcaactggc ttttagatttt ttatatcagc ctgtattctt tccctttgtt 12960
ttttagataa gttcctagca acagtacttg aagcttcac tgaccatata aaaacagaag 13020
caccctttac ggaagataat ctaaaccat cagcatggat tgaaacaaaa atatctgctc 13080
catatttctt agcatcctgg tatctattat tttaaactaa tgtctcatct ccatctctaa 13140
tcatcacggg tctgtaccca taagtatctc ttaaggttct ttctaactcc ttcgcaataa 13200
gtaacgtcac atctttttct aaaatattat ttgggccaac cgcaccaggg tatttaccac 13260
cgtgaccgc atctatagca acaacaatat ctctaatact tttattgagg tttttatttc 13320
ttttcacctt aagctctaatt tttatatattt ctgtatttat agtctgagtt gggttttgcc 13380
aatggactga ttcataataa tcaacaacga ttctggtaag acttccatcc tgagatgctc 13440
ttactttttt gattggatag ttgtatggaa catttatctc ggtcggaagg ctcgattgat 13500
taatttccat gacaattcta gagggatttt taaagaataa gacctacca atgaaacttt 13560
atctagacta aagttaatac taatttcatt gttacccatg tcttgtattt catogaagaa 13620
aacatcattt ccgctgataa aaaaagatag aaaaccaagg attgccaatc tattcataag 13680
tttttaacca ttctttaaag ccattgtcac cagaaattaa agaaacctct cttccttcag 13740
gaagatggct gaaaattatt tttagatcaa aacttctttc atgttgaagc ctttcaggcc 13800
actcaattag cactactttc ttagagttaa tttttctact tagatcaaat atatcaatat 13860
cttctgcttc gttagttcta taaagatcaa tatgcaaaaa cattaaatta ttaaaatcat 13920
attcctcgca gagagtgtaa gttggacttt ttacaagatc cttccatcca caatttttta 13980
taatagatct agatataaag gttttgcctg ctctaagtc tccctcaaga tgtatttcaa 14040
tttcttgcca agaagattta agtatctcca tagctatttt tgaccctagt tggttttag 14100
cctcgtcatt tataagagta agttttttca togatttatt aattgtctta ttataggfat 14160
taaactagag gcatttaagc ctatttcacc aatatctacc ttaaatttta aacctgcctc 14220
cgaatgaaca gcaacagcaa taatgctcgc atttcttata tcaagacctt gggcaagaag 14280

agcagtcaaa acaccagcca gaacgtcacc agttcctcct gaagcaagtt caggtccgcc 14340
gcacgcgcat ataaaagact tgttatttgt tgagtcaaag accaccgttc ccattccttt 14400
taatatgaca atagaagccc cataagtgtc tgcaatttgc ttggcagcag aaatcctatc 14460
tctttgaact tcttcaatag atatgtttag taatatagcg gcctctcctg gatggggcgt 14520
catcaagatt gttttattgg attttttgat taaagatttt gatgatgcca cgatgtgtaa 14580
agcacctgca tctaataata tttttgaatt atttgcagac tttaatattt taccoaatat 14640
agtttttgca aaagcagtat tagcaattcc tgggtccatac aatataacgc tatgatcttt 14700
aatttttaaa tctatatcga ccccttcaac catatcaaca ccaattgcca ttacctcagg 14760
gtttcttaat aaagatgggc ttacattaga cgtgtctgtg acgagagtta ctagtccaga 14820
gccacaaaat aaacttgcct cagatgcaag tattccggcg ccacccatgc ctggggagcc 14880
agcgcatatc aaaacttttc caaaattacc tttatggcta tcctttgctc tattgggcag 14940
aagatttttt aattcttgaa aggtaaaact ttgcaatata gacatgaaat tacttaagct 15000
aaattatggt gaatgctaaa aattcattct tcagcgaaat aaccaatcct atcggaatgg 15060
aatgatccca taaaaatttt ttcgtcatga ggcactgcaa tagtgggtag gccaaaaact 15120
gttctagaaa atgaaaactt atcttctctc ttaagtgttt ttgattaag tttatgaatg 15180
gagaagggta gagaacaatt tgttttttct gcacaattac caaatcatt agcctgtaaa 15240
tctaaagaag ttaaccaaac cgaaccgtct tttaaaaaga tattatctgg actctgaata 15300
aaataactac ctgttttagt attttgatta atatcgtaaa cagataaatt gtcaccttga 15360
ttatagttaa cataaagaag tccagagccc tcatctagca aaattccatt aggcccactt 15420
ccatctgttc catcaacttt attaaaatta ttgtcactcc aaagtactac atgccctgaa 15480
atacttttaa ataatgaatt cattaaccat tctttcatgc tgatatctct tttatacata 15540
tgagatgcat aaaaacttcc atctttttta agggcaacat cattgaaata atattggtca 15600
gggaccctaa tacagcccg ccatatcata tcccatgaag attcattttt tataatttca 15660
aacatttcaa tcgactcaaa tggcgaatga ttaattacag ctagttgata gaagccttca 15720
tcattttcaa caagatctat tccgtgggga ttgaatatgt ctagctcgcc tcttatgcat 15780
gaagcgtcac cccaagagct ttctccaaaa gtaatttttg gtaccttttt ttcataggta 15840

tttaaatacca ttaaggcaaaa gtaaccaggg gtatgttctg catatggacc gatccctcca 15900
aattcagaaa taaagaagaa tttattatca ggcgtaatca caatatcttc tgggtttgag 15960
aaattacaaa taaccttgat acggtcacg gattcacact tactaatatc catttggggg 16020
cctatatagt cagtagatac gacagttaca gaaataaata aataaagaat agaaaccggg 16080
acggtaatct tataatagtg ctttataaaa atttctaaaa tctttgaagc atgatttgga 16140
agtgcgatca gccaaacccc ttttaaaaaa gatagagccc ccataacaac aaaaattact 16200
gaccatagcc tgtcagacca ttccggctga agaagaccag ttaaaaaaag aatcatccca 16260
aagaataaag ctaaatagcc tgagattttg actttcgcac caacaaattt cgaaaagaga 16320
tatgtataaa gaggccttat aagaactagt aaaccacagg ccaagaaaaa gaatgctaag 16380
tagtaattca taagttagtt tttatataaa tgctccttaa taatactaac aagttctaag 16440
ggcttgtcca atggaacatg gtgagcagct ccaggaaccc cctcaaaagt cataatgtca 16500
ccatatgtat ttttaatat gtccaagata cttccggagg ttaataagct gtcttcaccg 16560
tggtatgaaca aagcagggca gccaaatgaa aatgtgtaac cgaataacct ttcaagactg 16620
ctaaacatga catcatcaaa tttccatctc caccagcct caatatTTTT tactgagtgc 16680
tcagcaatgt atcttaagta ccaatcattc gtacaatctt gcttaggcac taacctaaac 16740
cttttaataa tatctgtctt gtcttgatag tgcttgatca ttctgagagg agaagagtgt 16800
tgattcgggt cataatccgg tggctttata aatgtatcaa taataataat attatttatt 16860
agatcctttc tttcagatgc aacgtaacca gcaacatgtc cgccgagggg gtgtccaaca 16920
ataaaaatat ttgaaatatt ttttttatcc ttttcctttt caattacaga gacaatacat 16980
tctccaaaat ctttaatgcc atatgaatct ctaaaagaag agtcacccat gccaggaaga 17040
tctattgcaa ctatatattgc gcagtctcta aagtggggcg caataggatc ccaccatttt 17100
ttatgagcac ctgttccgtg aataagaatt attaaatctt tgctttcatc tttggagtgc 17160
cagctagaat aggatatatc cccatgagga ttcttgataa tctctgagct aggcttgtcc 17220
tcaatggcat ctttgaacca ctgtggggca tgaataatgt cttgatttag attgttagtt 17280
atttcataa acagtattct aagctataaa aaataaaaat atgaataaac ttaatttaac 17340
gccagcagca actgttttag tcttaaagga ttctcctgat gggatggaag ttttgatggt 17400

aaaaagatca agtaggcctc ccttcggaga cctttttgtt ttcccgggcg gcaagattga 17460
cgaaggatgat ttcaataata agatagaaga tttttgtgag ggcgtgactg ataaagagggc 17520
ctccataaat cttggattag attctggagg tctagcatat tgggttgcat gtattagaga 17580
atgctttgag gaggttggaa ttttacttgc taaaaaaaag agtggggaag atcttgatct 17640
agatggagtc gataaacata aatatcaaaa atatagagag atgttggtta ataatgaaat 17700
tgatttatat aaaatctgtt tagaagaaaa tttaattcta atgcctcaac aaatagcccc 17760
tttctcgcat tggataaccc ctgaaataga aactaggaga tttgatacac gtttttttat 17820
tgcccacctc cccaagcctc agaccggaga acatgatggt agtgagctca tagacagtgt 17880
ttggatttca ccaaaagaag cgctcaaaaa atctcggttcg ggtgagatgc ctatgattat 17940
gcctacaata aaaaatttgg aacaatgtgc acaatttgat tcgggctcta agcttttaga 18000
aaatcagagg aatctctcaa atgaggatat cccaccaatc ctgccaaagt tttttaaaga 18060
agatggtgag tggagggggtc tattgcctgg agataaaggg tatgaggatc attaaataat 18120
atggacttaa ttactaaaat aacagctccc aaccctggtg ttttacttgg ggggtgggact 18180
aatacttatt tgattggcaa agatgatata acccttgctg accctggtcc aaatatatct 18240
gagcatctag atgaaattat caaagcaggg gatgggaaaa taaaagaat ctttggttact 18300
catacccata cagatcattc cccagccgca ttgcctttat caaaaactct taatgttcca 18360
atgtacggaa ggctagtaga tggatgaatcc tcatgggagg atgaaacatt tatcccagat 18420
attattttta atgataaaga tattattgag acagacgaat atacgttaga agtaatacat 18480
actcctgggc acgcatctaa tcatttatgc tttttaataa aagatacgaa atgccttcta 18540
acaggcgatc acattatgga cgggtctacg gttgttattg ggccaccaga tggcaatatg 18600
acaagctata tcaattcatt agaaaagtta ctagattttg atattgattg ctttgcgctt 18660
gggcatggaa attatattca tgagcctgag aaaaccattc aatcaattat taggcacaga 18720
ctaacaagag aaagaaaagc tcttagaaag ctaggagagg caggaatctc atcattagat 18780
aaacttacta agcttggtta tgatgatgtg tcagagatgc tccatcctat agctaaatat 18840
agtctagaag cacatttatt aaagcttata gatgaaaaga aagttaaatt agataaagat 18900
ctattcgaaa taatttaatc ctttttattt ttatgtaaga ctttctcctc aatagcttct 18960

atatcaatat catcaattga gtcttcgtta ttatcaggta tctttttttac atcttttttca 19020
attttaaggt cgataggaga ggctccaaga tcaaaagtta gctcccttac attttttgaa 19080
atagtatcct caacacagtc atcttgatcg tatgcctcac gggctctctcc tttatcatta 19140
atgggaaagg gtctttgggg agggcccatt tgcattgcatt taatagtagc aactgggtgag 19200
tagtaatcgt cactagaata taacttatca agttcttttg gtgatattga gcaaccaata 19260
attccaaata gaataaatgg cgctaagagt cttttcataa actttgtttg ttttctataa 19320
gagactcaac aaccgaagga tcggcaaggg ttgtcgtgtc tcctaaatta gatagatcat 19380
tctcagcaat ctttctaaga attctgcgca taatttttcc cgatcttggt ttaggcaagc 19440
ctggagcatt ttgaattaaa tctggttttg caatagctcc aatttcttta gcaacaaatt 19500
gtttcaattc ataactaaag ttgtcatcaa atgattcatt tatcattaaa gtaacaaaag 19560
catatattcc ttgcccctta attggatgat caaaaccaac aacagctgct tcagcaattt 19620
tagggtaggag cacaagagca ctttcaattt cagctgtacc tagtctgtgg ccagaaacat 19680
taagaacatc atcaactctt ccggttatcc agaagtatcc atcctcgtct cgcctggctc 19740
catcaccagt aaagtaaata tctttataca taccaaaata ggtgtcgatc attctttgggt 19800
gatcaccata aatacttcta atttgactag gccaaagattg ctcaataact aaattacctg 19860
cattagagcc ttctagcgta tttccatgct cgtcatagag agatggctta actccgaaga 19920
agggcagagt tgctgacca ggttttggtg gagtaatacc cgctattgga gagataagta 19980
cagaaccagt ttcggtttgc caccaagtat caataacctc gcaattagat ttaccgacaa 20040
cactgtagta ccaatcccac gcctctggat taattggctc tccaactgta cctaaaattc 20100
ttaggctatc tctttttgtc tttttaacag gatcatcgcc ttgggccatc agagctctaa 20160
tggcagttgg agctgtataa aaaatactaa tgtcatgctt atcgcatatc tcccaacacc 20220
ttgatgctga agggtaggtt ggcactcctt caaacataag tgttggtgct ccatttgaaa 20280
gaggctccgta caagatataa gtatgtcctg ttatccatcc cacatctgca gtacaccagt 20340
atttgcctc tggccttatt ccaaaaagat atttgaaact aatatgagcg cctaatagat 20400
aacctgcagt agtatgtaga acacccttg gcttgctgt agagcctgat gtatagagaa 20460
tgaaaagagg gtcttcggaa tccataggct ctggagcaca cttattagaa acatctttaa 20520

caagatcttc ataccaaaca tcttttttat catcccaatt aatttcgcca ccagttcttt 20580
ttataaccag tgtattttta acatctggac agcccagaag agcctcatct acattagatt 20640
taagtggcac ttttttgccg ccccttaaac cttcatcagc agttataaca attttacaat 20700
cagcatcaag aattctatct ttgagtgatt ctggagaaaa gccaccaag acaacagagt 20760
gcacggcacc tattcgctg caagcaagca ttgcgaacgc agtctcaatg atcataggca 20820
tataaatata aactctcgag cccttttgaa caccaggtc ttttaaaaca ttagcgaact 20880
tacatacttc gtcatggagc tctttgtagg tcaattcttt agaatcagca gggtcatctc 20940
cttcccatat taatgctatt ttgtttggat cattttctaa atgcctatcg atgcagttta 21000
agctaataat tgttttacca ccctcaaacc acttcgcatt attaaattga ttattgaatg 21060
ttgttttgaa gtcttccatc cagcttatgt tttcgtttgc tagattttta aaaaatttag 21120
aaggatcttc tatggattgc ttgtaaagtt ctttgtattc atcaaagtct tttatataag 21180
gattacttga atgttttggg ctataaagct ttcgaggcat tcttaaataa ttgaagggtg 21240
ggggttaatg aaattcttcc ctttgggatt ggacattatt tttgtaatga gcgattcgta 21300
atcgctatca ttggtttcta aattaatata tgcagcatta attattaata aaggggcaga 21360
actatagtct aagaaaaacc ttgagtatgc atcattcagt ctttccaggt agtcaagagt 21420
tagatattgt tcgttaatat ttcctctctt agtaatcctg tcttttaaca catcaatagg 21480
tgctgaaga tagattacta ggtcgggtgt tggcgcgtcc aggggttagat ggtcatatac 21540
tttgtcatat agatccattt cctcgtaga aagagtaacc tcagcgaata atcgatcttt 21600
ttctattaaa aaatcagcaa ccctcactgt ttcaaaaagg cttctttggt taagatcttg 21660
aatttggtgc attctttgaa acaagaagaa aagctgagtg gctagagctg attggcttgg 21720
gtttttataa aaattcttta agaagggtt ctctgccggt tgttctaaaa aagaatcata 21780
attaaatggt tcggctatct tatttgctaa agtagttttc ccaacaccta tcggtccttc 21840
aattgcgata tatttttgaa gtggtacttc tttaatagct gggttcattt ataacgtata 21900
gtttttatcc tagattatct caactcgta tagaagccaa atatttttgc agattcttca 21960
tcactttttt cattcatctc agcaataatg gttcctctca atgagtcac attttttttg 22020
aaatcagtcc accacttccc aataccactc ttgtctccag aagctctctc tcttacaagg 22080

agaaaatcat atgcagctct aaaccttggg tgtcgaagag ttttatatgg ctggctacca 22140
attctgctat gaagttttaa ttgaagtacc caaatatcct taatatagct tgaaaatttt 22200
cttggtattg ctgtgatttt ttgttggtcg cgaagtacac catccataga tcggaaaaat 22260
tttctcacat taatttctcc attcttagaa cactttttca ataacgaggg ccataacaat 22320
gcagccataa gaaagcctgg tgtaattgac tgttggtttt taaccctatc atcagtattt 22380
cttagagcat gtgtcataac attgctggca aagtcattcc tacttggaac agaaaggata 22440
agatatttat ttaaagaaa tgagcagagt ttttcaaaat tcttttcacc cattccattc 22500
aaaaatattt tgcagaactc atcaaacat ctagcgtttg agatgcctga taaaagatgg 22560
cctttatcat agatggcatc ctttaaccaga ttatctatct taaaattgag ttgtttacta 22620
aatcttatag ctcttaagct tctcactgga tcttcttcga atcgtctttg gggatctcca 22680
atagatacaa taaccttctt gtgtatatgc ttaagcccat cgttatgac ttctattttt 22740
tttgtaacag gacagtagta aagggcattc acagtaaaat ctcttctatg acaatcttgt 22800
tcaagagtgc cccaattatt atctctaaga atctttcctg tcgaatcagt aacaatattt 22860
tcaccatctt cctgatcact tctgatctg aaagttgcta ctcaagcaa ttcactccta 22920
ttaaaaacat gaaccaattt aaatcttttg ccaattattc ttgaagcttt aaatgttttc 22980
ctgatctgct caggcggtggc attagtggct atatcgaaat cttttggctc tagtcccgtt 23040
agtgcacgc gaacgcaccc accaaccaga taggcctgaa aattattttt ttgaaggctt 23100
tgaacgacag atatagcaaa ttacttatc ttattattat ctatcaaatc taattatgaa 23160
tttttatgat ttaagggcat tagccattta gctgcttttc ttttaatttcg tcaagtgttt 23220
tgcaatgaat gcaatgagtt gcagttggtc tagcttcaag cctcttaatt ccaatttcat 23280
caccacagga ttgcaccaa ccataatcat cttgcttaat ttgttcaata gatagacct 23340
ttttgctgat aagttttctc tctctatctc tggttcttaa ttcaaaagca aactcttctc 23400
cttgagaagc cctgtctact gggctctgat aggtttcacc tttagctcta agatgatcaa 23460
aagttttttg catttcatcc ttaagatggt ctttccaaag aagaagaacg gcaacaaaat 23520
gtttcttcat tgctgcactc atatattttt cacccttctt agatttataa ggtgcaattt 23580
tagatttatt attagctatt gttgcttttg cagacttctt ggctgcaact tttttaactg 23640

gagctttttt ggtaacagtt tttttaactg tagctttttt ggcaacagtt tttttaactg 23700
gagctttttt tactgtcttg gatttttttt cgaccatgta agattttata gaattttagg 23760
gtggagaaaa tatcagatac tgacaaaatt agctagtcac tttttaattt atttaataact 23820
ttcaagtacc catccgagct aagtcttggc ccaaagtgtc caacaacctt ggaagatgca 23880
tagtttgcaa acttagcaca tgcttcaata ttattccctt gaaggtaggc atgcataaac 23940
gatccggcaa acatatcacc agctccattg gtatctattg gagttatttc ttctgcttga 24000
gcatgcttct caaccccttt atctataaca acacttccat cggcgctttt tgtaatagca 24060
gtcatatagg gcttttcttt ataaaagcta acagcatcat caaggctttc tttaccagaa 24120
aaagcaacag cttcatcatc attacagaag atcatatcta ttccatatga ctctattaaa 24180
tcaaattttt ctttaaaacc atgaacaata cctgcatcag aaagagacaa ggctttcttt 24240
acgtccttgt ctttaagggtg ctctaagact gaaataacag cattaaagtt atcgtcactt 24300
gttaccatgt agccttcgat ataaaaaatt tttgaatttt ctacaacatc aaaatctata 24360
tctgatttac caagatacgc actaactcca agcatgctgc tcatagttct cttagcgtca 24420
ggagtaacta aaattaagca tttcccagtt gggtgatctg tattttcaga gctgacacca 24480
atatgtttga ctccagccga cctgagacta tcaagatagt ttcttccatc ttcatcatca 24540
gaaactctgc atacatgatg gcaattcgaa ccataatttg ctgcagcaac aagagaattg 24600
gttgagagc caccgcaatc agaaatcgat tcggctccca tttcaataag tttgctaatt 24660
ataggtgcct gttcttcaga agatgaaaga gtcataagat cggctacaag gcctacactt 24720
gataaaaatt catggctgac tttatattga gtatctacta aagcatttcc aagggcgcta 24780
atatcatatt tcatgtgtta ttcttttttg tattatttgt tttactgttt ctaacgttct 24840
attaatttct tgatccttat gcattgctga aatgaagcct gcttcatatt tagagggggc 24900
aaagtatatt ccacttctaa tacatgaatt taaaaaattt gaaaatagta catcatcagt 24960
ttttgcaaca tcattaatat tattagggag ttcttctgaa aaaaagaatc cgaacattcc 25020
accaattctg tttatagaaa acggaatacc tgactcaatc attaagggtt tcatcccatc 25080
caaaagcaca gatgcatttc tttctaactc tttaaacgga ttttctttta tcaacaattg 25140
caacaaagca gttcctccag ccatagctag tggattgcc a gacaaagttc ctgcttgata 25200

gacaggacca gaaggagcta gatagttcat aattttcttct ttgcctccaa aagctccaac 25260
aggaagaccg ccacctatta ctttccccaa agcagttaag tcaggagtaa tattataaat 25320
ttcttgagct ccgcctagcg aaactctaaa gccgctcatg acctcatcaa atattaaaat 25380
agaattattg gctgaggtgg tttccctcaa taactttaag aaatcttcat gacctggaac 25440
aaagcccata ttccttgcta ctggttcgac tattacggcc gctaagtcac cttttatctc 25500
attaaatatt tctaaaaact gttctttatt attgtattcg caactaaatg tatattttgc 25560
caaactgca ggaaccctg gagagtcagg taagccaaag gtggcaacc cagaaccgc 25620
cttaattaaa agagagtcaa catgaccgtg ataacagcca tcaaatttaa taattttatt 25680
ccttcctgta aaacctctgg ccaatctaact cggtgtcatg gttgcttcag tacctgaatt 25740
aaccattctt attttttcaa ttgaaggaat gcatttttta attagcctgg ccacatcaga 25800
ttcaagactt gtcggggcgc cataactagt tccaagcgca acttgatttt taattgcacc 25860
tacaatgtct ggggtgtgaat gaccattat catcggaccc caagacccaa tataatcaat 25920
atattcatta tgatcagcat cataaaggta tgcgcgggat gctctttcaa aaaatatagg 25980
attgccattg atatttttga atgcccttac tgggtgaattt acccctccag gcataagagt 26040
cttggcctct ttgaataagg ctatggattt atcaatttta ttatgtgtca aagctaattc 26100
ttcttaattt ttaatttcga tatgatatca acttaattct tattatgtat ttaattgttt 26160
tgaaaattca tccacagtat tccagttggt gaactcataa gtatttgtaa catctgtagg 26220
accctttgta atccacataa tcaatctaact catgtgttta tctataaatt tatattttgg 26280
ataatctatt ttgcctgcaa agactgctaa tttcattggg ttccaaggag acagctctaa 26340
aaatttttgc atatacgggt ttgtttcagg tgtatttttt tcaggctttc tcgcaactac 26400
attaactgaa aagaaggcat tttcttttgt ttcaagacaa gcaacatttt tttgaataaa 26460
ttcataaagt tctggtttgt gtttgccata cctaattgctc gcaccaataa taattttatc 26520
aaattgatat aaatctaact ctactgcttt tgcaatatgt attatttttg aagattcaga 26580
cacatctaact ttagagaaaa tttttttaca aatttccaga gtttgcccat cggttgtaga 26640
gtagattagt agagttgatt tcataaatat attttatatt gttaattaaa aaatttcatg 26700
gcacgcgcaa aatataaaca aaaatatagt atatttaaaa aatgattaaa tattttattt 26760

ttatcgcttt attcatcagc aactttgcat atgcaaatat tgatgcagct aggtgcgcag 26820
ggattagttc agatcaagaa agacttgatt gctatgattt aatttttaaa gctaatgatg 26880
agctaccttt agatagtaat attaaaactc tcatcacacc agctatcaag gctgtgaccc 26940
cagctgattc aataaaaatt gaaaataaag caacaaaaga aaaggatttc gggcttccaa 27000
aaacaaaaat caaaaactct gcaaagaact caataaaaac ctcggtagta aggattaaaa 27060
aaacaaaaag tggtaaatta atttttactt tagaaaatga gcaagaatgg actgctgaaa 27120
cttcttatag agcaaggaat atgttttaac cagaaaccgc agtcatttta gaagaggccc 27180
tggtttagtg tttttatatg attaatataa gtaataaaca gaaaattaga ataaagaggt 27240
tgaaatagca atgaccatac agagtatagc aacgaccgag ggtgccatga aaagaataag 27300
atctgttctg ggggtcccaag atggctcaag cttcagggtt tatgttaccg gaggtgggtg 27360
ctctgggttt caatatggct tcaagtttga taacgatatt gcattcgatg atgatgttat 27420
aaattgtgga gatttctcgc ttttaataga ctctatgtcg taccctatc tttatggatc 27480
aactctggat tttgttgaag atctctcagg ggctaaattc gttattaaaa acccaaatgc 27540
caaaacaaca tgcgggtgtg gagagtcatt tacagtttag attttgtaat tgaacccaaa 27600
agtcctttta cattcttgct tgatgtctga actatcagag cctcattact gactctttta 27660
tatcccatcc aggcaaatgc catggactca atagcaaaaa catcatggcc taaatcactt 27720
gagaggacga tatcattact agccatctca gaaattcttt ttaccagata cttgttatgg 27780
gcaccaccgc cacaaatgac aatgtcacia ttattatgac cttttttatg gattgaattt 27840
attatagatt ttgctgaaaa ttcaaccagg gtacatagaa tatcttcagc tttctttttt 27900
aagaatttct tggataatat ttttatatta aagagctctt tgctgttga ttttggacat 27960
tttcttttaa aaaaattatt ttgaagcaat cttcctaact caatgtgatc tacttttctt 28020
ttggctgcaa ttgcaccatt tttatcaaaa ggaatctgaa gaaaatcact acaatatgca 28080
tctaaaattg cattaccagg gcctacatca gttccccata tgtcatttct attttttaca 28140
aatgagtaat ttgaaatccc cctatatatt aaaataattc gtgggtttct agctttataa 28200
aataattgat tatggaattc cggaacaaga ggggcgcctt cgccaccag agcaatatgc 28260
atattttctaa aatcacttac caccagaagg cccgtttctt ttgcaacaat atttggatca 28320

ccaatttgca tagaaaatgg aaatctctta ttaatttcat gcctaattgg ttgtccagaa 28380
atagcaacac actctataga tgatttctta attttactaa atccaatcgc ctcathtaata 28440
gattttgaaa ataagaatcc tatctcttta ttgatagtgc cttaaactga caacgagctt 28500
tcgttathtt caataagttt ttttactttt aatcttaggg attttgggaa ttttaattgaa 28560
tggaataact caagatgtat ttttgtacca atgctaagaa atgaaatata aatagcgtca 28620
tgactagtcc ctgtcattgc tccaatatag atctttttac tcattactac ttggagaaag 28680
tttattgaac tcttccattg ataatttatt attcttttaat agtgtatcaa aacttgccct 28740
taaatttttt ggtactgggt cagctgatgg taattttact tttataggat ctattctttt 28800
cttacctata ttaaactcat aatgtaaatg aggcctgtt gccagtcctg aactcccaac 28860
gaaaccaata gtatcgccct gagaaacttt tttgcctttt cttataccct tactaaattt 28920
ttctaaatga caatatctag tcgaatactc atttgtatga tttataacta tctcattacc 28980
acatccattc ctttgaccag aaaaagaaac aatcccatcg cctgtagttc tcaccggcga 29040
tctctttttt gctgcataat cgactccatt gtgagctctt attgtatgaa gaacaggatg 29100
cattctgttt ggattaaaat gagaactaat atatgcaaaa tctaaagggg ctcttaaaaa 29160
ggccttttgc atattatttc cattttcatc aaagtactgt tttttactgg cttcagtga 29220
aaacctattt gcaaaatagg tattaccatt gtttaataaac ttagcaatta cgatatcgcc 29280
attttttact ttctctccat cactataagg agtgtcataa attacatgaa attcgtcacc 29340
ctctctaata tcaaaaacga agtctacatc ccagccaaag atataagcaa agtccataat 29400
cacactttcg ggaatatattg catctagagc tgccctcataa aatgaggatt gaataatccc 29460
gctattgtaa gattcaatta actcaatact tttgettata ttcttaaata ctatttctgg 29520
agtaaggcta attgaaatag aatttatttg atctttcatg atctcaattc tgggttagctc 29580
ctcaccagaa tattcaaaga gcattttttc ccctgggtta atgttagcaa tgatattttt 29640
ggagtcta atctaaaaatt tataagcagt atttaaaggc actgaaaagt tttcaaaaat 29700
tattgaaaga ttttctccat cttgcacctc atgcatctga tatgttttcg tttgctcgag 29760
aggtatagag aatttttcag taatctcaat ttcttcaact ggcaaagatt catatgtttc 29820
tatatcaata taaagcatta atattaagac aatagaaaca gcaaaagaaa caagtaccgc 29880

tcttttttggga actttttttaa aacctatcat tagtcctctt gaatatTTaa gagttcagtg 29940
ttgccaccaa atgctaccga gttcttttgag acaacctttt cgTTaagaaa ttgcagaagg 30000
taattaggcc caccagcctt ataaccgctg ccggatagat tttgtccacc aaaaggctga 30060
gaaccaacaa cagctcccac catatctcta tttatgtaaa tatttccaac attacacttg 30120
tcactaaata tgtcggccct tttttctact ctagtatgaa taccattgtt aaggccgaag 30180
ccactatcat taatatTTtc aattagtga tccatttcat ttgatttaaa tctaacgata 30240
tgaagtattg gtccaaattg ctcgTctttt agatcagaaa tattatcaat ctcaattatt 30300
gttgagagcta caaaattttc atctacatta tcagcagacc taaaaataga ataattctta 30360
tcttcaaaac ctttgacgta agcattgaga ctatcaagag atgttttgct aattattgga 30420
ccgatatctg tgtcaagatt ttgagggttt ccaattttta gttctttcat accaccctta 30480
atcattgata aaagatcgTc atatatttTc tcttgTcac agagaactct caaggcagag 30540
catcgctgTc cagagctatc aaatgctgaa cgaataatat catcggtTgc ttgctcgaga 30600
agtgcgcttg aatcaacaat cattgaattt atgcctcctg tctctgcaat caaaggaata 30660
attgattcat ggttactTgc aaggctgctt tgaatttttt ttgctgtttt taaagaccg 30720
gtgaatgcaa cccctttgat gttattaacc tttgaaagca tgtcaccatg aatgccgtct 30780
cctaggatta aatttaaagc atcttttggg accccaatct catgaaattt gttcacaatg 30840
atgtagccaa gaattgaagt atgttctgag ggcttaactg ttactttatt tccacatgca 30900
agggctgcac ttatttgccc tataagtatt gcaacaggaa agttccatgg actaatacat 30960
aaaatatggc cttttggtga gtaagatagg gcattgatct cgccggtggg accctcaaga 31020
atatggtctt ctgtttggag gccaacagct tgTTTTgcat agtatctgag aaaatctata 31080
gcttctctaa tttcatcaat agtatTTTgt actgtttttc cggttTcatt cataaggtaa 31140
taaattagct cagaaggatt tgcttcaatg tcgtctgcaa ttttttctaa aatggatgct 31200
cttttttcaa catgcatcaa agaccattca ctaatatTTt gttcttcaag gctgcccttt 31260
atataatcta gatcatcata tgaggctgta ccaatatTTc ttccatctgc aagagatgag 31320
atgtcatgag tattggTTTT cttataatcc tttgctttat aaatcgatga agcctttatt 31380
tcttttgaat caaactttcc aagctcctct tcaagcattt ctaaattcac cctttcactt 31440

aaatcaaadc ctttagagtt cggctctgtca ttaaataatat ttcttggcat tggatatctct 31500
tttttttcat cctcaatttt taaatggggc cctctggcaa gccaaagcaga atctgttttca 31560
ggatcaagta acctattaat aaatgagctg ttgtctccat ttctaatag tcttcttaca 31620
agatagggtg gcaagtcttt atatttgcca atcggcgcac aaattgaggt atttttttca 31680
gtattttaaga ttttgtttgc ggacttataa agcagctctc ccattccaaa tagtctttga 31740
aattcataat ccttatttga gccaaagtga tggattgcag aaatgggatg tgcattatgt 31800
gtagcaaatt tagggtaaatt tttttccaca ttaaaaattc tttttgcaca ggctaaataa 31860
gctaagtctg taacagattt ttttgtatag acaggataac catcatagcc atagatttga 31920
gcatgcttaa tttcataatc ccaatatgct cctttttacga gccgaacatg cataggtgct 31980
ctgttttcta gtagctcttc taaccaatct attgtggcta tagctctttt gccatatgcc 32040
tgaacagcaa taccaaattc tttccagttt ttaatatttg gtgaaagggc catctcttta 32100
ataatctcta aactaacggc cagtctgtct tgcctctcag catcgattgt aatttctaca 32160
tcttttagact ttgcatactc tgtgagctga ataagttttg gaagtagatc agactttatg 32220
tcttttagct ttttcatttc atatctagga gataaagctg atatttttat tgagacacca 32280
ttaattgtat tttttgtcag atttatcttg ccgacttcat ctatagcatt cttataagac 32340
tgataataag tgcgggcttg ttcagcgttt cttgctgcct ctccaagcat gtcaaatgaa 32400
taaatttcat tttctatatt ctttaattttt tttatgtcat caaatctct gcccataaca 32460
aattcttgac tgagaatatg catggcacca actacagcat ttctaattgg aaactcgcca 32520
gattttgaga ttaaagaact taatagggcg cttgggtttt tgggtccactc atcggttgta 32580
gaaactacct tgctgcaag aagaagaccc catgttgatg cattaacgaa aacactatcc 32640
gctttgttta aatgctctat ccacgcaccc tcagataact tctcagatat tattagatcc 32700
cttgtttttt tgtctgggat tcttagtatt gattcggcta aacacatcag agcaacacct 32760
tctttattat ctagcccata ttcgcttaga aatgcatcta gttttgttct ctcgctctta 32820
ttctctctgc aagcatcaat tattttattg agcatttttt gaaatcgaag ggtcatttaa 32880
aaaatcagaa ttactaataa gatctgaaac tatttcttgt tcaggataaa acttgttact 32940
tgttaatgtc atagctatta ttttaattct atagtgccaa gtatcaactt actgattcta 33000

ttaatgactt aaaaacatta tgatagtcac atgagtgatg cacttaaatt aattaaacga 33060
ggaaccgacg agatcctcac agagtctgat ttaaaaaaga aattagattc tggaaaacag 33120
ctaatacatta aggcgggggtt tgatccaaca gctccagacc ttcatttggg tcataccgta 33180
ttattaaata agctaagaca ttttcaagat cttgggcata aagtaatttt tctaattggg 33240
gattttactg gtcaaattgg agatccctcc ggtaaaaata aaactaggcc aacacttact 33300
tcagaagaat taatttcaaa tgcaaaaaca tatgagaaac aagtttttaa aattcttaaa 33360
aaagaattaa cagaagttaa atttaattct gagtgggtgca acaagcttgg tgcagatggg 33420
ttgattgggc ttgcatcaaa atataatgtt gcaagaatgc tggagcgtga tgattttaat 33480
aaacgttata ggcgcaaatca aagcatagct attcatgaat ttttataccc ccttgttcaa 33540
ggatatgact ctgtagccct agaggctgat gtcgaatgcg gaggaacaga tcaaaaattt 33600
aatttgctag tagggagaga gctacaaaga tcctatgggc aagaacctca gggtgtttta 33660
actgtaccca ttctagaagg cttggacgga ataaataaaa tgtccaaatc attaaataac 33720
tttatagcaa tagatgaaga gcctaagat atgttcggta aaataatgtc tatttcagac 33780
gagttaatgt ggagatgggt tgagttactc agctttacct cagagaaaga aataggagtt 33840
cttaagaaga aaatggaaga agggaccaac ccaagagata ttaagtttct tttagcagaa 33900
gagttagtag atagattcca ttcagagggt gatgggtcga aatgcaagga agcttttctt 33960
caaagatttc aaaaagggtca aatgcctgat gacattcctt ccatgtcagt tgatgttggg 34020
gctgagggca ttccattagt aaacctttta aagaattgtg agatgacatc aagtacatct 34080
gaggcgatga gacttgtaa acaaggcgga gttaaaatcg actctgtaaa aatagaagat 34140
cctaaaatgc taatttcaaa aggccaggag tctatttatc aggttggcaa aagaaaattt 34200
ttaaaaatta aaacataatg aaaaataaat tagtccaaat atttcttctt ctttttatag 34260
ttgcgtgtaa tcaggatagc cctaataatta aacaaattag taatatgcag tattttattg 34320
ataatgaaat aagggaagga atctattctg ttgaaccagg cttgcaatac tcaattatcc 34380
aaaatggaga tcaaagttct gaatcaccat tgctgcagga tacaattaca gctcatttct 34440
acgggaccct cactgatggc tcagtttttt ggagttctgt tgaaatgggt gagcccttaa 34500
cagtcgaact atcaggtcta atagttgggt gccaaaaaat aatctctatg atgaagaaag 34560

gtgatgaatg gagagtttat atcgacccaa gtatggccta tggcgatgag ggccggcctg 34620
ggataccttc aaactcaatt cttatTTTTg atattgagtt attagatata caaaaaaact 34680
aacccttata tataacggat agagcgtata cataaacctc agcaacctgg ttaataatTT 34740
tatttttagg tgttccggct ccatgaccag ctcttccttc aattctaatt aatattggat 34800
tatcacaccc ctgggactct tgaagttttg ctgcaaactt gaatgaatga gagggacta 34860
ctctatcata tcttttagct gtagtaatca gagtagttgg atagcactca ccttcaacaa 34920
tattatgcag aggcgataa gctaataagt tttcaaactc atctttctta tctggagagc 34980
catagtcaat ttcccaggcc cagcctatag taaatttatg aaatctaagc atgtccaaga 35040
cgccaacttg tggaattgag acttttaaata aattaggatt ttgcaacatg gtcgcagcca 35100
ctagcaaacc tccattagag ccacctgaa tagcggttga tgatggagaa ccaatttctt 35160
gtgcgtgtaa aaatttagca gaatatgcaa agtcatcaaa aacattttgt ttattaaaaa 35220
gcctgectgc atcatgccag ttatcaccat attcaccacc acctcttaga ttaactacag 35280
caactattcc cccttggttc atccaagtga ggtagctttt actaaagcca ggtagtcttg 35340
agatattaaa gccaccatat ccatataaaa gtattggagt attactgtct atttttaagg 35400
actttttgta actaagatgg atgggtatTT gtgttccatc tttggatgga aagaatttaa 35460
agtcagatgt aaataaagtt gagtcgtgtc ctttaagatc ttctttccaa aaaagtctct 35520
gtgacatata agttagatta attttgtata tttctcgagg agttacaaaa tttgtaaatg 35580
aaaaataaga tacctcatct tcaatttcac caccaaaacc acccattggt ccttttcttt 35640
ctgttgctag cttatTTTTa tatgtcctt ttaaatacaa gaagtgaacc tcagtaaaag 35700
tatcaacgag ataagaaact acaatagaat tatttataaa gctaacta ctaatagaat 35760
tagtactttc accaacgact tcattccaaa caaatgagcc attttttatt gtaagtgata 35820
ctacttttcc gttcgcagca ttttcagttg agtaaaacca aaaagtatca ttcttgcttt 35880
ctaaaaagct ataagcgctt attagctcat ctattaaagg aataaaaggc tgatcagggc 35940
ttagctgaac atagagtcta tttctttcat ctgtgccttc accaatagat agaaacttaa 36000
ttttagaatc ttttacaaca cttattcccc aactccatct tggcttttca ggattctcat 36060
aaacaatcac atcctcatct tgagcagtag caattttatg aaacattaac tttggagctg 36120

tattaatatc ttttaaaagc tcttccgatg gctcgtcgta tttttggtaa taaaatccag 36180
aatcatcatt ttcccatgaa gcaccagaaa atttagccca ttcaatccta tcatcaagag 36240
tttttcctga ttcaatatca agtactttcc aggttctcca atcagaccca ccatctgata 36300
ttgagaaagc tagaagagat gcatcattac taacgctggc acttgcaaga gagatagttc 36360
catcttctga gaactgattt ggatcaagta aaactctatc ttggcattct tcacagtcct 36420
taatcatcag cttgctttgt tgccatgacc catcattgaa ataataaaaa gtttttttgt 36480
taacctgata aggcattgctt atcgaatctg tatcccaaac ctcatctaaa ttcttagcaa 36540
tagatTTTTT atatttattt tggcctataa attttttgtgt gaaattattt tgtctctcga 36600
cccagtctgt tgagtcctca cttgtaaaat cctccatcca tctataagca tcttcgatta 36660
ggatatccatg gacttcttca ctaaaaggaa ctttatttga ttcagggtat tcaaaattat 36720
ctgtttgatt tgagcaacta actagaatta atagagctgc catgtatatg tatttattgc 36780
gcataaaaag cataatactt tacaattacc ttatgaatca attcaatcga gcatggcaac 36840
ttttacgcaa aaaattggaa gttttttatt gtacttgccg caccatcat gatgcttgaa 36900
atagcaatgg catcaatggc aacccaata caaaacgtta ctcagcctga ggatattcta 36960
gaatTTTTTca atgagaatat agcctttcta ggctctgtta gtctcttagg cgtagtttta 37020
agtatggctt ttatgggagc gctttttgtt tcgtatgcat ccatagaatc agaaaatgag 37080
attgagcctc taaatgcctt attttttaggc ataagaaaat tctttccact tctgggggct 37140
tatcttatag cttcagttgg tgtctttttt ggtatcttat tattaatact cccagctttt 37200
tatgtagcag caaggctttg catttttcct gcatttatta tgcttgaaga caaaggagct 37260
atagaatctc ttaaattatc ttgggaaaag acggatgagc atggcaccac tttgtttggg 37320
cttaccatta ccttcttttc ttttaacaatg atttttgcat cagttgcca atccattata 37380
agtccgggat taatgcaatt agttgttctt gcaattattg aatatgtaat agtaattcca 37440
tggggctatg tatatTTTtag tttatacaag tcattaaaaa gatattaggc agattaacta 37500
gaataaatta atacgaagct tttcgtttac aaaaaaacac accactctta taatccttec 37560
tccgaagatt tgcactttgt taattcatgc aaacaacttt gggctctgtag ctcagcttgg 37620
ttagagcgca cccctgataa ggggtgaggc ggtggttcga gtccaccag acccaccatt 37680

.ctttgaacat taacttttat ttccctttata cttattctta actaaataag gatagtatta 37740
tggctaaagg tttagacaag caaaaaaacg acaagaagaa aggcaaaact ctaaaagaaa 37800
aaagagcggc aaaaaaagaa aaaaagaaat agttttttgt agaagactat tattatctag 37860
tcgacctagg gccatcttta aaatctatct taaagaacaa cattattccg gtcactgagc 37920
ttactaaagc taaaatagaa aatattttta aaagaagggt gtcgatatta tctctttctt 37980
tccaatccat aatatgaaac ccccatca gatcccatat ttccattta ttagatctaa 38040
tggccacaac ctctgctgag taaacattaa tgtaaacatt taattcagta cctttaacat 38100
ttttactttt tactctataa attggcaatg atctgcctct atattcagat cctgattttt 38160
catttacaac ttcttcggta gcaaatggca aaagagttgt ttggcttgaa actgaatcta 38220
tagcatcttc catagaaatc tttgttaaag gctgcccaag catattcaga tacttagttg 38280
agccctttgt tgtgatgata acaatctctt gacctaatct ttttttgaat ttaacttctt 38340
tagctttctc tatttcaaaa ttttaacttac ttaaatacaa agatgtttca acatgattaa 38400
ggtattgctc gcctcttacg agctcaatct tattaaaagc aaagtatatg cctgagatag 38460
tccaaagtaa taactgaaga gagataaaaa aacttaggta cttatgaatt ttccctaacta 38520
aaaagttcat cttttcataa tcttcattat ttcattcaatt ttacatcaa attcgtcatc 38580
accttttaaa gattccttaa tgcattcacg caaatgggtt tttagaattt tgccctcaac 38640
agtaataatt gaattcttaa gtgcttttat ttgattaaga atatcaacgc agtattgacc 38700
ttcctcaacc attccttgaa cgcctcgac ctggccctct attcttttta agcttattaa 38760
ttgctcttta tgacatggat gtttcatttt ataaaaacct cggaactata tacgtaaaaa 38820
ttgtagcgaa taaaaaatt gctacagatg catagtacag atacctagat ttttttctag 38880
tttctgaaca tgctctgcct agttctgggt cagcagggca aggcatatga taagttttgt 38940
agttgatata tcctgcaatg acaatcatta ctaatgcaa tagtgttatg taaagcttat 39000
attgagatag ggtaattaag aatggaaata cagttaccag gcttgcaaaa cttgccccag 39060
caccaagagc acaaaaaata gctggcaagg cacaacatat caaagtagaa gaagatgcaa 39120
ataatgaaaa aaaattcgac gccttatcat tcatagttat taaatttcta ggatttgaag 39180
atcagacata ttttggccat tagaaagaat cttcttttcg atctcagaga aggtgatttc 39240

ttttgcctta gaaaaggcaa tcactacttt tccattatca agatcaacat ctattctttt 39300
gacatcttta tcttttagaa aagtttttga aatacctctt gcacagaaat cacaaaccat 39360
tcccttcacg cttacgatcg ctatatattac atttgagagt ccttcaacaa atgcatcaaa 39420
cctctcaggg ttacatcaa gatcttttcc atcaaccatg gtttcatgta aatgaccttc 39480
atgagaatgc ccatgcatac cactatgggtc attctccgcc gcaataaggc cgcttgaaag 39540
caaagtaatt aataataatt ttttcatatt tttctccttt taaaatctgt acataaaatg 39600
taaatcccag ctcgatatctg cgtcatagcc aaactcaaca agggcattac cataaaaaaa 39660
tttcaattct ggatacgtag atgatttatt tgttacagta ttttttttgg ttttaaccat 39720
aaccatgta tgtaaatcat tgtagtcagc aacgtaagga gcaaatacaa gttgaatata 39780
ttcttcttta tagtctttgc taaatttact gttaatatcc ttttaaccaa agcccgcata 39840
ccacttccgg gtttcccagt ctccatggat accataaaaa ttatttttaa tacttccagg 39900
cacaattcct gattgaaaat ataaattcct ttgggatgta aaagtatttt ttctattcac 39960
caaataagta aaacgaaaat agttttctgt gttcttataa aattttatttt tagccctttc 40020
taggccaatg gaatatttgt aggttggaga gtagtggaaa taaatagaat cattaaaggt 40080
atctgattta tacattatgg ttgttccgcc agagtaggat atcgggtctg catcagcatt 40140
aaagaccatt aaaaaaacta aaatgattat tttatatata cccatggggg gtatcttacc 40200
ataggattat aaattaattt caggcatctc tccaagcgcc cattttcaat gaaatcacca 40260
tcagaattcc aaaaataacc atagctatag cagatgcata aaaaatacta actagcccat 40320
atgagagcca gtaaacagct accaatgcta taacagaaaa tactataagc ctgatggccg 40380
tgattataat tggccatttc attgcattgg ctccctgaga tgcaaaatac agggaaagcc 40440
catatccttg aaatacataa catacaccaa ggatttgaat atactgcttt gtaactaata 40500
aggtctctgg atccgatgta aaaatactaa tccaaagatt tgggtgttagt gctaaagcta 40560
gaccgataac agccgaaagc aatccagctg ttgttgcacc aaccatgcct attttttctg 40620
ctcgctctat atttttggca cctatatattg ttccaaccat ggctgtcata gcagtcccaa 40680
tgccgaacac aattgggatt aataagaact ctacccttga accaattcca taaccggcaa 40740
tagcagacgt tccaaattga ccaatgagtc ctgtcagtaa aagaactgtt gcaactgtca 40800

ttattggcga aagagatgcg ggtaaggcaa cagaaaaaat atcttcgaat aattctcttt 40860
caattgtttag tcttttgagc ttaagcctta caggcgatga agggctcgaa atttttatta 40920
aagttactaa tgccataaaa cctgaagtaa ctatcattga ccaagcagat cccacaagcc 40980
caagtttagg tagaccgaat gatcctaata taaatccagc agagaaaaaa acttgtatac 41040
ctgcacaaat taccgttagg actgcaggaa attgcatatc tcccatccct cttagggcgg 41100
cggttaagct acctgatagc caaacaacta ttgctcccaa taaatagaca aaacaataag 41160
ctaaagactc ctctaataat gcacctgttc cacctaatat tcttaataga ggctctccaa 41220
aaataaaaaa gataattaaa aaagctaaag ctccaaagca agatatatac aaagaatgcc 41280
ataataattt ttcagcccta ggcttgtctg cggctccaag acttcttgca atagaagatg 41340
tgaccgctcc tccaagagca ccaaatgcca tctgctgagt cagcatgatt gcaggaaatg 41400
ccaaagtaac agcagcta at ggagttatgc caagttgact aataaaccag aattcagcaa 41460
gaacaaccac agcatttatt aaaaatgcaa cagtattagg tgctgacatt ttaatcagca 41520
aaggaaatat tggatctttt aaaaattgct ctgttctctt gtccataacc ttctcattta 41580
ttaatattta tttggttaaa gggtgaaaga gaacacgttc tcaggattca cgaatcttta 41640
aataataata acagctttta gttaatttaa tttttaaaaa caaccaaate tattatctgc 41700
tagattttat tttttaaaça tattaatate taatttaatt gcaaaggggt ggccatttcg 41760
gcctgagatc cttggaaact ctttgaacct gatccataca ataatggcgg aggaattgca 41820
tgaaaatcat taaaaatcat ttattttatt caataatctt actatatttt tttagtctag 41880
agacgagctc tcaagctata gaagaagtta ttatcaaagg agactggaga gaaactagtc 41940
tgtcagcaga agactcaagc attgcagtc ttgatagtaa attaatagaa tcccaggctt 42000
taaagcactt tgaaaacctt tcgtacctag taccaaattt aaattttgct gcaagtgatt 42060
ctagagcaag acatttccag ataagaggaa ttggagaaag atctggctat gaaagaactc 42120
caaactctgc agtaggttta ctgattgatg atatagactt ttctgggcaa ggtggaattg 42180
ccacaacttt tgatgttgat caaattgaag ttcacagagg cccccaagga gcaagaatag 42240
ggctctagcgc aatggcgggg ctcatatata tttctacaaa agaccaaca gaaagctttg 42300
aagggaaggg tgaaatagtt atgggggtcat atggaacttt taataccgga attgctgttg 42360

gcggctcctgt aaattttaat aaagacctca cttatagggt agctataaag aaagattatt 42420
ctgacgggatt taggaagaat atttttctca acaagtctga tacttctaag aaagatgaga 42480
gcacatttag attaaaagtg aattgggtaa ctgataacca aaccacttat aagtttctta 42540
tatctcaaat agagtttagat gatcctgcag atatatggac tatagatggg agtcttaata 42600
cattatctga taagcctgga atggactctc aaaaaagtaa tgcctatggg gtaaaaattt 42660
atcatcaatt taaaaaattt gaatttcaaa gcctatcaag cttaacgaat actgatgtca 42720
ttcttagtta tgatgctgat tggggcaacc ctgagtcgca ttcacccttt atttatgact 42780
atttttcaga aactacaaga aaaagagata ctttttagtca agaatttaga cttgtatccc 42840
aatttgcaga taaaaataca gaaaaaagca tcgaatgggt cgttggggct agttttgtag 42900
atataaatga aacaaatgct aaaaaagata ctggtattta tggagatcca tcagacccat 42960
atggtcctta ttttagtaac tcttcttctt tgagtgactt ttcttcttca agctattctt 43020
tatttggaat tattgattat ttaattaatg aaacaataaa aatttcaatt ggtggaagat 43080
gggagaattt taaatctaata tattttgatt cttatgacga atcattttca ccatcaaata 43140
aaatgtctgg tggttaagttg tcaactagtta aaactcttaa taataattct aatatttatt 43200
ttaatattgc taagggctat aaccaggggt gatttaattt aggtcttggt cttgataaaa 43260
attcatcaaa tagaaattta tattatgatc cagaattttt aactaattac gaagttggaa 43320
ttaatagtaa atttttccag tcaaaaattaa atcttggagc agtcctggtt tattctgatc 43380
gaaaagatca acaggtctta atttcaaccc aggttgatcc ttcagatccc aacacttttt 43440
tatacttaac ccaaaatgct gcagaaggaa ttaataatgg cttagagtta aatatagatt 43500
atgcactgaa taaatctcta ggtatatattg ctaattttgg attgctaaat acagaaataa 43560
aaaattggat ttcaagacca gatatagaag gtagagaaca ggcgcátgct ccaaaaaata 43620
gtttttcaat aggcattcaat tggaagccaa caaaccaatc ttatttatca ttgaatgttg 43680
ttggtaaaag tgagttctat tactctgatt ccataacaa tacttctgag tcatacaact 43740
taacaaacat taattatgga tatgaacatg gacaatggac ttattcatta tgggcaagaa 43800
atatttttga taagtactat tcagtaagag gtttttactt tggtaacgag gcacctgatt 43860
ttatagatac gctctacaga agacatggag atccaaggca tataggagtc atggtccaat 43920

atgattttcta actttataag tgaatttttgg atggaaatag cagctgttgt atttgctatt 43980
atztatttat tacttgctgt aaaacaagac gtaaagtgtt ggtttgcggc tataattagc 44040
tcaataattat atttctttat tatgtatgac gcaggcttat atatggaagc ttacctgcaa 44100
attttttata tcatgatggc cttttatggg cttcaacagt ggagaagtgt cgatactgat 44160
gctcctcaat ttattgttag aacatgggat aaaaggatgc atattaagat cattacatta 44220
atagttatca tgactttaat ctctggattt ttattagaaa aatataccaa tgctatcttg 44280
cctttcatag atggactaac aacatgggga gcaatagtgt cgacatatat ggttgccaaa 44340
agacttcttg aaaattggat atattggttt gttatagatt ttatttctat ctttttattc 44400
atgtccagag gacttttatt aacctcagga ttattcttta tttatcttgt aataatatat 44460
tttggttaca tgtcatgggt aaaaataaga gacgatatta gtgcagaatc ttcataacaa 44520
tctagatata aaaaatcata atctagagat tattaaaaca attaaatcag gccctgtatc 44580
tgaaatatct atttgtaatt ttgataacat caaagcaatc ttaagagttg atcatccatg 44640
cgcacacaaa ataaatgtag atcgagaaaa tgaaattttt atactcagcc aactaaaaat 44700
tttagatttt agtcagagg ttttatttag tgatttgtct tatggaattt tagtatggag 44760
atatatcgag ggtattgaat tttcacttgg taaggattcc aatgaagttt ttttaaaaac 44820
acttgggaacc gaattaaaaa aaattcatga tattgatctc cccaaaagta agaaaaaata 44880
ttttagcaat gacataaatt tttacagaaa tttactaaag gaggttcctg aaaatataat 44940
tcttcataga ggatttgatt tatacgacaa actcaataac tctgataatt atgttctctc 45000
tcataatgac ttgaataaga caaaccttct ttggagggat aggttatttt ttttagactg 45060
ggagtactcg agttttaaca atcctttttt tgatattgcc tcgttatcga atgcttataa 45120
cttatcaaaa gttgataggg caattttatg gaaagcttat acaaataatg aatattcagt 45180
attaaatgat acaaatctta gagaatggat gcatttttgt cattatttag agtacatgtg 45240
gagtatttcg ctcatacaaa atggaaaaat tgatcagaac accctaaatt taaaaaaatt 45300
agagaaaaaa ttaaaaaata ttatttaaac aacaagtatt gcgcattctt tagttgttgt 45360
tttactatta tgtaagaaat ttacatttaa atggggggtt gatatgaaaa ttttatgtgt 45420
cttatatgat gatccaaaaa cagggtatgcc agaaaggtat gcaagagatg atttaccaaa 45480

gttagataag tatcctgatg gaatgacact tccatcccca aaatctatag attttactcc 45540
tggtgagtta cttggttggtg tatctggaga actagggctt cgaaagtttc ttgaagatgc 45600
tggccataca ctagttgtta cttctgataa ggatggagat ggatgtgagg ctgataaaga 45660
attagtagat gctgatattg ttatatcaca acctttcttc ccatattatt taacaagaga 45720
caagatgaaa acagcgcta atttaaaaat ggcaattaca gctggcattg ggtctgatca 45780
tgttgatctt caggcagcca tggataatag cgtggatgtt gttgaagtta cttactgtaa 45840
ttctcgttca gtcgccgaac acattgtgat gatgaccta tcgatgggtc gtgattatca 45900
taccaacat agaattgtta aagagggagg atggaatata gctgatgctg tacaagggtc 45960
atatgatgta gagggcatgc atgttggtac tgtagctgca ggtagaattg gtattgatat 46020
gctaagaaaa atgaaacctt ttgatgtgca ttgacattac ttgatattc ataaactctc 46080
tgatgaaata gaagcagaac taaacctcac ctatcatgat tctgtagagt cattagttgc 46140
tgtgtgtgat gtagttaata ttagttgccc attgcatcct aaaactgagc acttattcga 46200
tgatgaaatg attagtaaaa tgaaaagagg tgcatatatc atcaatactg ctctgggcaa 46260
gatttgtgat aaagatgcta ttgcaagagg cttagagtca ggccagctaa gtggttatgc 46320
tggtgatgtt tggttcccac aaccagctcc aaatgatcac gtatggagaa caatgcctaa 46380
ccacggcatg actcctcata cttcaggtac ttctgtatct gctcaaacia gatatgccgc 46440
tggggttaga gaaattctag aatgttattt tgcaggcgaa ccaattagag acccatattt 46500
gattgttcaa aatggatgat ttgcaggtat ggggtgcgcac tcgtacacia aggggtacagc 46560
cacagatggc tcagaagagg ccgctaagta taaaaaataa gttttagaac ttacttagcc 46620
tttccttaat aatagagtct gcctctgcca taatgctatg cattaattct tccactgtcg 46680
ggatgtcatt aaccagtcca gcaaccattc cacatgacca ggctccaacc tccatagtcc 46740
cttcatgcat aatttttgga tagactcctg caacctcatc cacaatatca gcaaaagtta 46800
attcatcacc aagagctttt tctttttcaa tcaatctctc aacagcttca ttattaagaa 46860
ccctttctgt atttggttaat gatctcatga tgagtctagt atctaactca gaagcattta 46920
cgatagcctc ttccacattt tgatgaacag gtgcatcttg agtagcaata aacctagtcc 46980
ccatattcat tccctcagca cctagtgaca tggcagcaac caaacttctt ccatctgcca 47040

ttccacccga ggcaacaaat ggtatttcaa gctcatctgc tgctctaggt aagagtatga 47100
aattaggaat gtcattcttct cctgggtgtc cgccacactc aaaaccatct acagaaaccg 47160
catcgcaacc tattgcttgt gcttttaagt agtgccttac agagggtgcat ttatgaataa 47220
cttttattcc tgcttctttt aaagctggga ggtattctgc aggatttctt ccagcagtct 47280
ctacaacagg aaccctcgca tcaataatta cttttatcaa accaggatag tctgggggtg 47340
ttagtgatgg taaaaatggt aaattaacag caaatggctt attgggtcatt tctttgcctc 47400
tggcaatttc atttgctaatt ttctcaggcg taccctgtgt tagaccagta attgttccaa 47460
gtccacccgc atttgatact gccgctgcaa gctcggcaaa accaaccatgg tgcattccac 47520
cttgaatgat tggatgttct ataccataa attcagttat tttagttttc ataatactct 47580
cctatttctc ttccattggg gttaaaaaat cttcatattg agtttgaagt ctttgcctgc 47640
cacttatcca tctatctctg tcattacctt ttcttttaac atattctaga acctcactat 47700
gaggggttac taaaaacctc tcctcctcta tggcggttaag aacatccttt gcaacaatat 47760
cagcttccat cattccatca actcctgcta caccaggacc atttgcagtc atagcagttc 47820
taactgcttg agggcataaa caggaaacgc cgataccttt atttccatat gtaattttta 47880
tccattcagc aaaactcaca gctgcagctt ttgtaactgc atatccagcc gcacctagtt 47940
gagttaaaag accggctgct gaggatgtat tcataagata gccttcacct tgttctatca 48000
tttgaggaag cacatgtttt gcagcatgaa tatgagactg aacatttaca ccccatatca 48060
tgtcccaatc cgaagtatct gcttcaaaaa atcctggctt tccaccgata cctgcatttg 48120
aacaaaatat atcaatacca ccagaaaatt cattagcctt ttgtataaca ttataaatgt 48180
cgttttcttt ggaaacatca gcacttactg caagcccat aacactcttg gctgtttcct 48240
ccgccccatt taaattcata tcaacacata ctatagaact tgccccagat gcatagaact 48300
cttcacataa agccttacca attccactgg cagcccctgt aaccacaact cttttattat 48360
taattttcat aaattgacca cttttttaat tttttattta actttattgc ttagatagtt 48420
tattacaaaa caacaatact taatattgca aaatcgcata cccatcttta taaatttatt 48480
gctattatta gtgctagaaa atgaaatact gagattttaat atgaagaatg ttgttggtat 48540
tggtcgtcc ggggcaatag gaaaagcctt tattgatagc tatatcaaag atgatgatgt 48600

tgaaaatata ttttcatttt caagaacagg cctttccatt gaggataaaa aactccatag 48660
ttttttttatt gatattgagg atgaaactag tatttgtgat gccgcagaga agatagacaa 48720
gtcctcaata gatgaaatta tcgtcgcaag tggaatactt cataataaag attttgggcc 48780
agaaaaaagt attagagatt taaatgcaga taacctttta aaggtcatta aggttaatac 48840
tateggccca acaattgttg gaaagtattt cattccattg cttaaataaaa aagaaaaaag 48900
cgtcttagca tttttaagtg caagagtcgg cagcatttct gataataaaa cagggtggtg 48960
gtatgcctat agagcgagta aaactgcact taatcaaact attaaaagtt ttagtattga 49020
attaagaaga accaatccaa atgccattat ttttgggtcta cagccaggaa cagtagatag 49080
tgaattaagc gaacctttta aaagaaatgt aaaagaaggt aatttattta ctccagaata 49140
tagtgtattg cagctaaaaa atattattga tacagcaagt ccatctgatt caggcaaact 49200
aatttcttgg gatggggaag agattcagcc atagttggtt atgaatatat tttcatatta 49260
gaataaattt ttaggggaaa aaatgagtat taaatattat gactggtcta aatttcaggc 49320
caacactagg ccaaataaag ttgccataag agagctagat aataacaaga tctatactta 49380
cggagaattg gataaaagat catcaaggct tgcacacat ctccaaagtt caggaataaa 49440
aaaaggagat cgtattgcga tactatcgct aaattgttca gaattttttg agctagaatt 49500
tgcttgcgga aagattgggg caatagagat accattaaat tggagattaa caaaaccgga 49560
gctcagttat attcttaatg atagtgaacc aaaaactcta atttatgaca atcagtttga 49620
agaaatggtg aaagagctaa aagaagaatg taatatttct gaaatcatag ctcttgatca 49680
atttgaccaa gaaagtgatt atgaaaaagt tttgagtaat gcttcaggca tttattatca 49740
ggaagaagtt gatctagaag ataacattat gattatgtat acctctggaa caaccggtca 49800
ccctaagggc gccatgatca cgcacaaaat gcagcttttt aatgttatta atttaggtat 49860
ttcagcagct gtttccctg aatcagtcca tttagttgtc cttcctttat ttcatacagg 49920
cggaatgaat tgttattcaa atccaattct tcatgcaggg ggcgagttaa tattacttaa 49980
agagtttgag cctgggaaag ttctatcaat tateggcagc tctgactatg gagttactca 50040
tctgtttgca gttccagccc cttatcaatt tatgatgaat catccagatt ttgaatcaac 50100
aaatttatca ggagttaagt atgctggagt tgggggcgca ccttgtgcag aggtatttt 50160

gaagacttat ataagcaagg gtgttttcgat gcagcaagga tgggggtatga cagaaactag 50220
tccagggtgct actgggtcttg aatcgtccga ggctgaaaga aaaataggat ctgctggaaa 50280
accgcttctt catactgagg tcaagggtggg tggagatgat ggggaatgaac tgcctgctgg 50340
agaagtaggc gagatttata tttaaaggccc aaatattaca cctggctatt ggaagaaaga 50400
agaggctact agagattctt ttgaagacgg gtgggttaaaa acagggtgatg ctgcttactt 50460
tgatgacgag ggtttttttat acatagttga tcgatggaag gatatgtata tctcagggtgg 50520
agaaaatggt tatccagctg aagtttaaaa tgttatctat cagttaccac aaatcgcaga 50580
agttggagtt attgggtattg atagccctaa gtgggggtgag actggtaaag cctttggtgc 50640
tttaaagccc gatcatgaat tgacggcaga agaagtcata gatcattggt taaaaaatct 50700
agcaaagtac aaaattccag agaaagttga gtttattgca gctcttccaa gaaatgctac 50760
aggtaaagtt ttaaaaagaa cattaagaga tatgtaatat ttttaagcca aaaaaaacc 50820
agcctaagct ggggtttttta tttaactatt taaagttatc taatgttaag cattagaaga 50880
ttctttaaca gcaacattcc atataattaa accaaataga atcttggtta caaagtcagc 50940
aaggttatag ataaggttta agttaagagc tgatccaccg tcacccatca ggtaacctgt 51000
gaaataacct acaggataaa tcgcccaccc aaagatgata atatacatca ttgtgttgta 51060
agctgattgc acagcaggac ttgcagtatt acatgcagat tttccttctc cagcccataa 51120
ttcataaatc atgtataccc aagctaaaca cccaataatg aatgcaggcc atgcagccat 51180
gattcctgct tcacccatgt aaccaaacac aagcataaca agagaaccaa ctagtaattt 51240
cttaaataat gatccagcaa cattagttgc agcagcaaga attaagtaga attcacatat 51300
taatagagga actgttagta accaatcaat gtatctaaat acagttggcg aatcaccagt 51360
ttcaatccat acccctctca tgtacatgta atgccagaaa gcaataccag taacaagacc 51420
agatacagtt aatgatgttt tccattttgc agaaactcta tctctttcaa caaagaaaaa 51480
tacagtagat gctaataaag cagcagtaac taacaaaaaa gaaacaccag tgtaatcact 51540
agcatcaagg tcaccaccac ctgcagcaaa tgtaggaagt gcaataacac tacctaata 51600
cagtaataat ttcatatata actcctatat taatgacaaa gaagcataat tgcctcccc 51660
ctaattaaat atgaaattta atagtaacca atattaatag aggttacaaa taaaaagcat 51720

ctatTTTTTT aatagaaaat atgtgtaata aaacaccata ctttaaatta atatttatTT 51780
tgcataagta gacttgTTta gtatcatatt taacàtgaaa gttgcaattt atcctggTtc 51840
ctttgacccc atcacaaatg gtcacacgga tattattgat aggggttgcg gactctttga 51900
caagggtggtt gttgcaatag ctaagagtga atcgaaaaac cctctTTTTa gcctagagga 51960
tagaattaat ttagcccaat ctatTTTTta aggaaatgaa aaagtagagg ttgttggttt 52020
tccaagaaag ttaacagttg atcttgcaaa agactatgga gcttgTgcaa ttataagagg 52080
cctacgagca gtttctgatt ttgaatatga atttcagtta gcaacaatga ataggTcgct 52140
ggctcccaat attgaaagca tttttTTaac accaaaagaa agtctcattt atgtatcttc 52200
tagcttaatt aaagaaatat cagacttaaa aggcgatata tcgaagtttg ttcacTctat 52260
agtagagcag gcacttcgag cgagtagaca cttagctctg acaagcttca caaaaaaag 52320
tagctctttg attgacgatc gttttacaaa tagtgctttt gcattttattg cagggtcgc 52380
cttctcttcc atatacattt aatttaaatt taaaatatcc tggactacca tcagctgagt 52440
agaagtcttt taatgtagtt cctccaacct ctattgcttg ctctagtatc ttttttccag 52500
cagctactaa tctttttacaa gcatctaaat ctaactcatt ggcatttttt agcggatgaa 52560
tttttagctaa gaaaaggctt tcagacgcat aaatattgcc gataccgact acattttttt 52620
gattcattaa atagctTTta atatttgTTt tagagtgact gcacTctgaa aaaaaatcct 52680
ttgcattaa atttttagaa agaggTtcag gcccaagggtt ttttattaac ttatgtttat 52740
cgatatcaga agtaagatgc attgatccaa accttctggg atcattataa ataattcttt 52800
cctcatcaaa aattaattca atatgatcat gtttgataaa gaagttttca ttattttttg 52860
caattctaag actccagac attcctaaat gaagaataat ttttttatca ttagatagtt 52920
taaaaattat atattttgct cttcttttcta agctttcgac aacttgatct tttactgagg 52980
tttcaaagct atccaccact ttccatctaa ggtttctggt gtgtatcctt gcttctttca 53040
gcaaagagcc tttaaatttt ttaattgctc tcagggttgt ttcaacctct ggaagttcag 53100
gcatttttaa ctgagtAAat tattaatttt tactatgtca gctggagtaa ttgttcccga 53160
tgcgagccct agtcttagat tggcaagtat gtaatcatat tttgcattag caagattttt 53220
ttccgcgctg tataagtttt tttctgcctg caagagatca acaacgtttc ttgttccaac 53280

tctgtagcca acttgagtcg cttccagggc actagtggcc gaaatcactg cttgtttttg 53340
agcatttaca ttgcaacta atgttaaaac atttgaaaac tgggatctga cttcttgaat 53400
aatccttctt tctgtaaata gagtattttc atttgctctt tcatactgtg aatatgcttg 53460
cttccttctt gagttaacgg cgccgccttg aaagagtggc atacttagct gaattgcata 53520
attccttctt cctgttactg atggaactgg aataccttgg ccattgatat taaaaccttc 53580
atagttaaata tggtttggtt cagattctga ctgacttcca acaatgtcta tcttaggtaa 53640
atgatttgaa gctacacttc ttgcactgct tttcgctgct ttctttctca aatatgctgc 53700
ttttaactgg tagttatttt ccaatgctaa ttcaacccat gtctcttttg aacttggtgt 53760
tggcaggtca ataagcaaac catctcccaa ttcatTTaag ctgaatattt ctctaccaat 53820
cagagcattt aaagactctc ttgcagaata aagtgatcct tctgttctaa ttcttgaggc 53880
tttacttaga tcaaagtcca attgagcctc ttgaactcca gttatggctg ataaccaac 53940
atcgaatctt tgttttgctt gatcaagttg ttttttaata gctttttctt cagatattgc 54000
tgcatttaga ttatcaatag ctctaagtac gccaaaataa agctcagcag ttcttactaa 54060
aagattttgc tgetcaaatg caaagtctgc ttcagcagca tctgtaagag atttagattg 54120
cctatattga aaccatgtat ctagtctaaa gagcggctga gtaaccctgg cagatgtaga 54180
aaaagagtta tattgctgct gcagttcttt gttttgatag tattcgttcc agttagttga 54240
tccactcaag gtaatactcg ggagaagcgc agctcttctt tgaaccttaa gctctttatc 54300
tgctaaatat gaatattccg ctgctttata tgtagggtca ttctcaagcg cttcattata 54360
gatatccaaa agactttcag atgaaatatt aaatgagata aaaagtgcta aaaatgattt 54420
tgtgtaaatt ttcataactt attttaacct attaataatg tttgcagtgc aaacatttat 54480
tttttttaat ttatttgaat ttattttata tttotttagag tagaatatct ctaacaagtt 54540
caataatttt ttataataaa aaggctttta aaattggcta aaaattcata tgacgctcag 54600
gcaattgaag tcttatctgg actagatcct gtcaaaaaaa gacctgggat gtatacggat 54660
acatctaacc caaatcattt aattcaagaa gttcttgata attcggttga tgaagctctt 54720
tcaggttatt gttcaaacat aaaagtatct gttctaaaag atggctttat taaggctctt 54780
gatgatggaa ggggtatgcc aattgatgag cacccggaac ataaagtttc aggtgttgag 54840

ctcatccttt gtaaacttca tgctggagcg aaattctctg gagatgatta taatttctct 54900
gggtggccttc atggtggttg agtttctgtt gtaaattgcc tatctgatga attagagggtt 54960
agagtaaaaa gagattctaa agaataccaa atcactttta ataattggaga taagtcttct 55020
gaattaaagc caattgggga agtggggctc agaaattctg ggacatcaat taaattcaaa 55080
ccaaatccta catattttga aactatagag attcagataa aacagcttaa gcatttatta 55140
aaggccaaag cagttctctg tcctggatta acgatagagt tcgttaatga aaaaaagact 55200
gatgataaac aaaagtggta ttttgaagat gggctcaaaa gctatttgat tgattcttcc 55260
gagggagcag acttggtttt gctagattca attgtatgct ctaaaaaatc tcaagctcaa 55320
gagcttgaat ttgcaatcaa ttggtcatta agaccccaa aaaataaact cgatgaaacc 55380
tatgtgaatc tcataccaac tgctcagggg ggctcacatt taaatggctt taaggctggg 55440
cttttagatt cattaagaaga attttgtgaa tacagaaatc tattgcctaa aggtttaaaa 55500
attaatgcag atgatgttct taataatgca atttttataa tttcatctaa gcttcagaat 55560
cctcaatttg cagggcaaac caaggaaaga ctagattcaa aagatcacat gtcgttcgctc 55620
tcaagtacca caaaagacat ttttaagtatt tggcttaaca ctcatacaga agagggcgaa 55680
agaatagcag aacttgcaat tatgtctgct cagacgagag caaaagtctc caatatagtt 55740
gaaagaaaga aaacttttag aggccagcc ttacctggaa aactttcgga ctgtaatagt 55800
caggacttaa atgaaacaga gcttttttta gttgaggggg actcagccgg agggctcgca 55860
aaacaagcaa gagaaagatc tttccaggca atcatgcctt tgagagggaa gattttaaat 55920
acttgggact tagaaagtgc agaaataata aaatctcagg agataaaaaa cctatcaact 55980
gcaattgggg ttctgccagg aaataatgac ctttcatcac taagatacgg aaaaatttgt 56040
attcttgcag atgctgattc agatggtctg catattgcaa ctttactttg tgcattgttt 56100
ctaaggcatt ataaatcttt agttcaagag ggaaggatat atatttcaat gcctcctcta 56160
tatagaatcg attctggtaa agatgttcta tatgcacttg atgataaaca gcgagatgaa 56220
atagttactg aatttaaaaa gaagaagggc aagcctaaag taaacattca aaggtttaaa 56280
ggacttggtg aaatgaatcc acctcaacta agagagactg tgatggacct tgctactcgt 56340
cagcttggtc agctttctat cagctcaagc gataatgcaa attctatgat ggacttactt 56400

ttgtccaaaa agaacgcacc agcaagaaaa gaatggcttg aaaagaaagg gtctctagca 56460
aaaatataaa tatgaaagaa caaataacct caattagcct caagcaatat gctgaagagt 56520
cttatcttaa ctatgcaatg tatgtcattt tagatagagc tttgcctaata attggagatg 56580
gccttaagcc tgttcaaaga agaatactct atgcaatgtc agagcttggg cttgatgctg 56640
gctcaaagta caaaaaatca gcaagaactg ttggagatgt tataggaaaa tttcatcccc 56700
atggagacgg cgctgcatat gaagctatgg ttttaatggc tcaaaaatttc tcattcaaat 56760
acccttttgt agatggtcaa ggtaactggg gttctcagga tgatccaaaa tcttttgctg 56820
caatgaggta tacagaatct aagttaacta aatttgcaaa tcttttaatc tctgaattga 56880
agtctggaac agtcgattgg cagcctaatt ttgatggctc tcttttagag ccagtaattt 56940
ttccagccaa actcccatct attttattga atggcacttc tggaattgct gtaggaatgg 57000
caacagatat tccatctcat aatattaatg aaattattga tgccacagta catcttattg 57060
ataatccaaa atcacagttg gttgatttac tcaagataat taatggtcct gattttctcaa 57120
ataattcgcc aataattgct agcaaagatg agctgaatga aatttattcg actggaaaag 57180
gcggttcaa agctcaagcc caatgggcgc aggataagaa tcaaatcatt atcaacgcat 57240
taccttatca agcatctggg tctaaaattt tagagcaaat agctgatcaa atgcttaaga 57300
aaaaaattcc aatggtggtt gatcttactg atgaaggaga ccacaaggag ccagtaaggc 57360
ttgtcataac tttaaaatcg aacagagtaa atgctgaaga tgtaatgaat cacctttttg 57420
catcaactga tttacaaaaa aattatagag taaatatgaa tttgatttca ttgaaagggtg 57480
gaccaaagt tttctcccta gttgatttat tgaaagaatg gctagtcttt agaaaagaaa 57540
ctgtaataag aaaactagaa catagactcg accaggtaaa cgataggctg catatccttg 57600
aggggttatt aattgtttat ttagatttag ataaagtaat aaagattatt agagaatcag 57660
atgagccaaa aaaagatatt attacagctt ttaaactttc tgatatccag gcaaatgcaa 57720
tccttgagat taggttaagg caactagcca agctagaaca aatcaagtta gaacaagaaa 57780
gagatacttt ggtagcagaa caaacagata ttgaaaaaat actaagctct aagacaagac 57840
ttaaaacatt aattaaaaat gagcttattg aaataaaaga tgagtgttga gaagtaagam 57900
aatctccaat aagagaagca acagaagcca aagttttttc tgaagaagaa actcttgtca 57960

ctgagcctat tacagtagtc ttgtctgcag ctggctggat cagaagtgca aaaggccatg 58020
agatagaccc cagctcactt tcctatagag gagaagatgt acttcaagat tatggaagag 58080
gaaagagcaa tcaagtttca gtttttcttg actcaaattg gaaggcttac tcacttgcaa 58140
gtcactctct tccatctgct agaggaatgg gtgatcctat tacaggaagg gtatctgcag 58200
attctggagt aaagttcatt tcttcattga ttgggaatga tgaagataaa ttcattgatta 58260
tgaatactgc tggatatggc tatatttcag agtttaaaaa tatggtttct aataagaaat 58320
caggaaaagc atttatgaaa atccctcatg aagcagacct tcttaaagcc attaaagtaa 58380
gagacgatca tttgtatata gcagcagttt caaatattgg cagactttta atttttaaga 58440
ttgatgaatt gccaaactctt ggaaagggca aagggaataa aataataaat ataccaaccg 58500
ctaaatztat agcaaaaagaa gagttaatga cccatgcgca acttgtttct gaggctagct 58560
ctttaaggat tgaaagcgga aagagattcc tcactttaaa actcaaagat ctagaaaact 58620
atatttctac aagagcaaaa aggggaaata tgcttcaca aggatatagg aaggtagata 58680
aatgattga agaggttgag ttagaagtta aagaagactg attatagatt tctcaaaaat 58740
ttttaatccc tcatcaacta gctcattttc tattattagg ctggagaaa acctaacagt 58800
agatgcattc gccttaagaa tcattaaccc attatgatgg gattttttta ttaaattcatc 58860
tatttgaatt ttgctatcct tgcttacttc cactccgacc cataaaccag cagaagttat 58920
tttttcaaaa catttggtgct tctcatttat ttattcaac aaattgataa atctaacttc 58980
tttctttaaa accttattta aaaatgattt tttagaaatg gtatctataa cttcattacc 59040
tatggcacat gcaattgggc ctctccaaa agtcgtgcca tgagaccctg ccgacatatg 59100
ttttgaaact ttatctgaag taagaatgcc tcctagtggg aagccattag atataccttt 59160
agcaaaacaa agaatatcag gagtaatgtt aaattgctca taagcaaaga gtgttccagt 59220
cctgccaatc ccggattgaa cttcatcaat tatgacaagt gccttatgtt tttttgcaag 59280
tttctttatt ttgctataa attttttatc tgcttttgtt attccagatt gccattgaac 59340
cagttctaata ataaccgctg cagtcttata tgaaaaaacc ttttctagat tagtaatgtc 59400
attatatgga tgatttttta tccctctagg cagaggcgca aaaccgtctg taagatgctt 59460
agcttttagct aaagctatac caagcatagt tcttccatga aatgacgttg aaaatgaaat 59520

gacctcattt ttattttttat taactgttga tgagcaaaat tttcttgcta tottaactgc 59580
agcctcaatt gattctgccc ctgaattgca aaaaaatact ttatctgcaa aagaattctt 59640
gcataacttt cttgccaaag taactgaggg ctcatthtatg tagagatttg ataaatgcca 59700
aagctcctca gattgctttt taagaatttt tattagatct ttatttgaat gaccaagggtt 59760
agtgcaggca atgcctgctg tgaaatcaat atattttctta ttatttaggt ccatacatg 59820
cgatccactc gctttttttta caacaaaatc tgcaggagcg taaaaaggca ccatataatt 59880
agtatattct ttaattatct ctttcataaa tatattgtat cgtgggtttca gttcaaaaat 59940
taaagacca aaaattttcta gtatctttta aaccaaacag ctactcatt ggctttaata 60000
gaataatatt tatatctagt atatctattg tatgcggtgg catagcattg atatttttct 60060
tttttgagac aacgctcatt ttgccttttg ctgggtttaga gcttggtatt ctattcactg 60120
cattctattt aagtttttaa tggagtgata aaaaagaaaa aatatttatt tctcaagatc 60180
ttgtaactat agaaaaaggc tctaattatg ctgaatataa atgggaagag ttcagggtcat 60240
ttacctcttt ccagggtttca aaagacagaa gagatcttct taagctaage tttagggtcta 60300
aaggcgagga tgttgagggtt ggcagctttc taaatgaaga tgataaaaaat gtattaatag 60360
aagagctaac tcagatcata gatacattaa atcacgattc cttctcaaag ccagagcttt 60420
aatttctttt taacttcttt tagctttaat tcttttattt cagggatgcc atttctaaaa 60480
tttggaatg attctccctg tattaatggc ataagatatg caattgcttt tggggtaacg 60540
tccattccat ttttagctat aaaagatttt ggtaattttt tttctaaatt agctattttt 60600
gccagtggag ctggttcaat tttccaccta tatttttttag ctttacctct aactattatg 60660
ggcataacac cattcatgcc ttctactgca tactgaactg ccttggtccc aacagccata 60720
gcttgctcta ggtotgtttt ggaggcaata tgtcgtgcac ttctttgtag atagtcagaa 60780
acagcccaat gatttttttag ttttaattta tcagtaatta agttggcaat gtaaggcgca 60840
actccaccaa gttgagcatg accaaatgca tctattgttg ctgattcaga aagaaatctg 60900
ttcttattat ttttcaacc ttgggatacc acaacaacac agtagccatt ttttttaaca 60960
acacttttaa cttctgctag aaattttttt tgattgaagg ttatttcagg taataaaaatt 61020
atatgaggtg catctccttt ttcttttctt gcaagagaag atgcagccgc catccagcca 61080

gcatgcctgc ccataacttc taatataaaa acttttgctg atgttgctgc cattgaccga 61140
acatctaatag aggccttctag agttgatggt gctatatatt tagccgccga cccaaatcca 61200
ggacaacagt ctgttaatac caaatcggtt tcaacagttt ttgggatggc tatacagtta 61260
attggataat ttaatttttt acttatctgc gaaactttaa atgcagtatc agctgaatca 61320
ttaccgccat tgtaaaaaaa atatcctata ttgtgcgctt taaaaacatc aataagcctt 61380
ttgtactctc tttegcctga ctcaagatct tttaacttaa acctacatga cccaaaagcc 61440
ccaccaggcc tatatttcaa tgactctaata gcagatattg attcttttga tgtatctatt 61500
agttcctcat tcagcgcgcc caatattcca ttcttgcttg cataaatttt accgatgtct 61560
ttatgcttct tggcctctaa aatcagagcg ccagctgtag catttataac tgcggtaacg 61620
ccgccggatt gagcataaaa cgcatttttt ttcatttttt ctocatcaaa tgtactgaaa 61680
taaaaatgat taaagtatta taactaatat gcgaatacat atcttaggga tctgtggrac 61740
ctttatgggc ggccttgcta agatacttaa agagtcaggg catgaaatat ctggatcaga 61800
tattcaattt tatcctccca tgtcagacta tcttgatagc tttagatattg aaatgatcaa 61860
gggctatgac ataaaaagca tgccagatgc tgacttgat gtgattggta atgcacttct 61920
tagaggaaat gaatccgttg agcatatttt gtccaattct cttccgttta agtcaggccc 61980
tgaaatgctt ggagaaattt taaaaataa aaagggttta gctatatctg gaacacatgg 62040
taaaacaacg acttcatata tgttgactca tattatgtta gatcagggtta gggatgtcgg 62100
ttttctagtt ggtggaatat caaataatat atctggctct gcatgtcttg ggtcagatgg 62160
aacttttctg attgaggcag atgaatatga ttctgcattt tttgataaaa gatcaaaatt 62220
tattcactat tcaccaagca ctatagtcac caataatata gaatttgatc atgctgatat 62280
ttttaataat ttagatgaca ttaaaagaca attccatcat ctaattaaaa taattccatc 62340
aaatggaaat gttgtttatt ttgctgatga taaaaatata agagatctta tcgatatggg 62400
aatttggta aatcagatag caatcaataa taatgctcat tcaattgagg cagtttatct 62460
tgataagact ctaaaatatg aagaaagtat ttattcatta aatgagttac ctttaataagg 62520
agaacataat tttaaaaact acatttcggc tattttggcg gcaaagacag atggaattca 62580
aattcaggat tctatcaaat cattagctag ctttgatggc gtaaaaagaa gattagaata 62640

taaaggaagt tttgatggca tagaaattta tgatgatttt gctcatcatc caaccgcaat 62700
agaattttct tctaatagcc tcgtaacaca aaatccatca aaaaaaatac ttggtctcat 62760
tgagctaggc tcaaacacta tgtctggggg ttctcatggc ttgtcttttg ttgaatctgc 62820
aaaatcttta actcatgtta tctggctaga tcgcaataat gttttgtcag agaatgctag 62880
cattgaatct actaacacta ctgaagattt tatttctgca gcgatatctg ctttctcaga 62940
ttatgatatt gttattttga tgaccaataa agacagccaa aaaatattaa aaccattgt 63000
agatcacttt gaaaaataat aatttaccag tttttccttt aggaatagtc gccctcccag 63060
gtagcatcca atctcttcaa atttttgagc ctaggtatat acagatgggtt aaaacatgtc 63120
tatccaagaa ccatggattt gtaattgttt ttaatgccaa taatgagtct caaggcgatt 63180
tcactttttc taagaaggga agttttgttg aaattataga ctttaataat ttgccaaatg 63240
gccttcttgg gataactgta aaatctataa ataaggatgat aatcagtaat atatgtcaat 63300
tagaagatgg actgcatatt gctgatatta aggcacagat agatccagag gtagatgatc 63360
aagctgtttt ggcagaatat cctgagatat ctagcattct ttctcagctt gtaaagcatc 63420
ctaagattag tgacctgcc atccagggtg actttggctc tgctgattca gttgcatacc 63480
acttagcagg ctttatacct ctaagctcaa atgagaaaca aaaactatta gaagcattcg 63540
atgcagcaca gcggatgaga attctttcag actatattga aagaatatct actacataaa 63600
ttattttatt taatattatt ggcggcttat tagattttgc tctaagccaa ttgatggact 63660
taaaaaccgt agggatttct acaaaccttt tttcaatata gtatttgcct ggatagttac 63720
taagcattag cccaataata atagtgaaca aacctggcca ggtaacacca gcatcatgat 63780
tcccccaata ataagggaat aacctaaaat atttttaatc agtaaaacta aataccacaa 63840
caacgggtta ttggtcttaa atttcgagac atctttcttt ataaaatagt cacttggaat 63900
taatccagca agccacctca tactaacgag actaaataaa aaaataaata aagatattga 63960
gctcagccaa agtataagaa ctttataaga ttcaaaaaaa attatgatgg catttaggg 64020
attaaattcc ataattacc acctttttat ttaataatga tttagtatag caatgtctaa 64080
aaataaattt aacattaaa taaatataca ataattttgc caaaaaaagc cttaaatgat 64140
atttgcactg aaacctcag tgataataaa gccgaaaatg ttctatcact tgatatcaaa 64200

ggatatttctt cttttgctga caacatcatt attgcaactg caaattcgaa taggcatgca 64260
aagtctttat ccgaaaagtt agttgaggag ataaaagcta ataaaattag tatcatgggt 64320
gtagagggca agacagaatc aggttgata ttagtagact gtggtgaggt tgtagtaaat 64380
attatgaaga atgacataag agagttttat gatttagaag gtctttgggg tgaaaacacc 64440
ctcatcgatt cttcgaaata aatgctatta aatatcataa gtggttgaaa ctcaccttca 64500
tcttgggagt taacaggaat agaattattac acaaagcaa tccctaaaga agtaagtcta 64560
aattttatta acgtaaaagg gcagcaacac ccaaaaagat ctacagaaga ggtcttaaaa 64620
cttgaatcta agttaatctc gtccaaaata gattctaatt gatataattgt ttgctgggac 64680
tcaagtggcg agtcgttgaa taattttgaa ttttagtaaat tttttgaaaa atccatgctg 64740
gaaaatatga agctttactt tgtaattgggt gggttcatttg gcataacctg agatattctt 64800
gataaatcta ataagataat ctcatctca agtctcacc tacctcacag gcttttcaag 64860
attgttctta tagagcaaat ctataggtca ttttcaattc tcaaaaatct tccctaccac 64920
aaatgattga tctgaatgag aggtttgttg aaaaaagaag ctttttcaat agattgttgc 64980
ttatatattt tttctttggg gccctgtttt tattcttgct tttcaagacc tattctttac 65040
aggtttctag ctattctgac tatgaactag cagcactaaa gaataaaaca aaagaagtgc 65100
tagttcagcc tgttcgagga gtcactatg acaggaatgg aaatattcta gttaataatg 65160
tccctacata tgacctata atccaaccat caaaaattaa aaacttagat gatttcatta 65220
ttaatatttc taaaataata gagctatcag attcagaaat tgaaaacatt attaaaaatt 65280
ttaaaaggag tgcaacttat aatagagagc taacaattaa aaaagatctc tcgaaagaag 65340
aaatagcaaa atttgaagta agaagctacc aatttcagaa tgcatttata gatgttaggt 65400
atagtagaga aaataaatat ccttacttat tttcacacgc actcggttat gtaggtggtg 65460
taagtaatga taaggtttta tcaattttaa aaaatcaaga tttgaagcaa tctgaaacaa 65520
cttttaagta ttcaggtggt tttattgctg gaaaaacagg attggaaaat atttatgatt 65580
cagctttaag agggtccttt ggaaaaaagt tatttgaagt tgatgcaagg ggaaggcttt 65640
taaaagagct aagttttgaa aaacctatta atggaaaaag cttttttact catttagatt 65700
taaattcaca aaaaaaagcc tttgaacaaa tgaataatag aaggggtgct gttgttgctt 65760

tagagcttaa atctggttct atagttacct acctaagcac tccaagcttt aatgtaaattg 65820
gtcttttctaa tggcatgtcg tcggtagagt tcgaaaaact gattaatgat gtagacaagc 65880
cattttttga tagggccggt caaggctcgtt actctccagc ttctacaata aagccagcaa 65940
tagcattggt tggattataa gaaaaaatag tagactggaa ctttacactt aaagatcctg 66000
gattttttgt attaccagag gatcagagga tttatagggg atggaagaaa ggagggtcatg 66060
gaacaattga tatgaagggt gcaatcatag aaagttctaa tacttttttc ttttctcttg 66120
cttataaggc tgatattaat aatttaataca gccatctttc tgagtttggt tttggcagga 66180
atgtttgtaa agattgtttt aatccagatt ctggattatt gcctacgcct gaatggaaaa 66240
tgaataatct taattttggc tggtttaag gagataccgt taatttaggt gttggtcagg 66300
gctatatgag tgccactcca gttcaattag catattactc tgcatttctg gcaaaaaaag 66360
gaaatcttca agagctatca tttgtcgaga gtgacagcct aagcaatact gcttttataa 66420
ataatttaag catagataac tcagactggg atcaaattca ctcaagcatg attggagtta 66480
ttgaagatcc gagaggcact gctaaaagat taaaaccttt aaaatcatat gttgttgctg 66540
caaaatctgg aacggttgag cttgtcagca cacaacaaa ggaagattac aaactagtaa 66600
ggcaaaatat tggtaacaga gaccatgcaa tcattgtggc ttttggtcca atgcctgac 66660
cagagtatgc agtaagtgtt gttattgaaa atggtgaaag cggcggttct gttgcaggtc 66720
ctgttgctat tgctgtttta aatgagctta tcaataaatg aagaaaaaat tagactttaa 66780
aaacttttagc atttattttg atcaatattt atttattgag ataaccctgt tgcgggtcat 66840
gggtttatatt tttttataca gcgcataca agaagatatc agcactgttg ctaagcaagc 66900
tgtattcgtt ggttttggtt tgctgttaat gtttgtagtt agccaacctg accctgattt 66960
ttataataca ttttctgggt tatttttttg ggggggastt gtattgattt ttctaactat 67020
gatttttggt aaagaaataa atggagccaa aagatggctt gatttaggat tttttaccct 67080
gcaatcctct gagattatta aaatatcatt gccaattttt ttatcatcat atttatataa 67140
taagccactt ccaataagca ctaggcatac ttttattaca ttgatattaa taggtttfat 67200
atttgcactt gtagctagac aacctgactt gggaaccagt ttagtagtat ttatgtcagg 67260
gggttatgta ttatttttag caggattaag ccgccgttta ttggatctgc aatagcctta 67320

tttttattat ccttacctbt tttatggaat aacttttctag agccatttca acaacagaga 67380
gttttaactt atttgaaccc agacgcagac ccttatggta ctgcatggaa tataactcaa 67440
tcaaaaatag caattggatc aggcggaatt aatggaaaag gttatcagga gggctcccaa 67500
gcccattcttg atdttttacc agagacagaa acagatttta tttttgctgt tattgctgaa 67560
gagtttggtt ttattggagt ctgtattttg ttatcagtat ttdtctttat atkactcaga 67620
tgtttatatt tagcatttaa tgcaagagat agattttgca ggtaactat aggaggccta 67680
agtttagttt ttgcctctac attatttatt aatttagcaa tggttggttg ttagtttctt 67740
gtagttggta tgcctcttcc atttatcagc aaaggcggct catctttgct atcctkttat 67800
atagcttttg ggattataat ttctatggca acacataaaa aattaatgca aagatgaaaa 67860
aaattatatt tataacttta atattcacta tttcaattac tgcggattat tcgaatcatg 67920
aagatagcca aatggtaata aatgaactcg tcacaaaaca cggttttgag gaatcttatg 67980
ttactgcaat cttaaaaaat gcaaaaaagc gtgatgagat gcttaaactt gttgctaata 68040
ctgctgaaaa aacaaaaaca tgggatgaat atagggttat ctttataaaa acaaaaagag 68100
tttcagaagg caaaaaattt ataaaaaaaa atattaatgc tttagagagg gctgagaaag 68160
aatttggagt tcctaaagaa ataattactg ctattttagg cgtagaaact aattatggca 68220
gcaataaagg tggatacaga gttttagata gtcttactac tttaggtttt gatgaccgcg 68280
gaagatctaa tttttttaga agagaactta tagagttctt ctttttaaca agagaaaata 68340
atttagatat caaaacaaca aaaggctcat atgcggggagc catgggatat gcacaattta 68400
tttcgtcaag ctaccgagcc tatgctgtag attttgatga agatgggttac gttgatttat 68460
ttaattctgt tgatgatgca attgggagta ttgcaaatta tctttatggt catggatgga 68520
agagagaagg aaagatcgta acaaaaactt atccaaacaa tgtagaaaa ttttataaac 68580
ctcatgagtc tctaacaagg ttcatacctt taatctttta tgaagatgga aaagatcttt 68640
tttttatttg tgatgataac tttagggtta ttgctaagta taatattagc gatgtctatg 68700
caatggcggt ttattactta tcagaagagt taaaaaatg aaaaattat tattcacatt 68760
actatcaact tcaatattta ttcaagcaca gagttttggt ccagattctc ctgagttaga 68820
tctgaagagc tatatcctaa ttgagccaaa caccaatact gttattgcag aatttaattc 68880

agatttggaa atagaaccag ctagcatgac taaaattatg actagctatg ttgtggctga 68940
tcagattgca aatgggtttaa tatctcttga tgacgaagta ctaattagtg aaaaagcatg 69000
gagaatgcaa gggctctaaga cctttataga agctggaaaa aaagttacgg tatctgatct 69060
tttaaaaggt attatgattc agtcaggcaa tgatgcctct gttgcaattg cagaatatgc 69120
aggcggcact gaaagaggct tcgttgactt aatgaactct tatgctgcct ctttggagat 69180
gaataatacg atctttcaga attcaacagg gcttccagat gaaaatcatt tttcatcagc 69240
aaaagattta gccaatctga cagctaacta cattaataaa tttccagaag aatatgcttt 69300
atataagcag aagcaattta cttttaacaa tattaagcaa ttaaatagaa ataagctttt 69360
atggagagac gactcagcag atgggtgtaa aactgggtcat acagaagctg ctggatactg 69420
tttagttggc tcagcaaaaa ggggcggtat gaggcttatt acagttgttg caggaagtaa 69480
gtcagataat gatcgttttt tatcttccca aagattactt gagtatggat ttagattttt 69540
tacaacacag aagatgtaa gtgcagaaaa agaatatcag aatattacag tttgggggtg 69600
acaagaaaag atacttggcg tgggagttct agaggatata tctattactc ttcctagaac 69660
aagtttcaaa aatgttgaaa ctgtttataa agttaataac aatatccagg ctccatcgt 69720
agttggtcaa aaagtcggga ctctagaaat aattagtaat gatgagattg ttctagttac 69780
ggatttgggt gcttttaaaaa atattgaagc taaaggtttt tttggaagaa tatgggtcaa 69840
gttcgttctt tggatatatta gcttatttgg tctaacagat gaaaacacca cttaaaggcg 69900
tttttaatgg ggcttttgat actgtagata acataaaaat atcacctttt tctcgagctt 69960
atactttttc tgatagcgta tatgaagtag ttcctttctt taattcaact gcaattgctt 70020
ttgatgatca tataaaaaga ttagaatttt ctgctagtca actagcgatg gccgtagatt 70080
taaaagaagt tatatttgaa attaatctt taataacatc atctgaattt agtaatggct 70140
atgtctatta tcaagttact cgaggcggtg atccattaag gtctcatatt catgagccaa 70200
atttaaaaat agaaaccttt ggctatgcaa aggcgcactt attccaatgg aaaccattaa 70260
gagtatcagt atgtgatgat attaggtggg ggaggtgtga tattaaatca acatcattgc 70320
ttggtaatgt tatgaacatg aatgcagcaa aactagataa ttgtgatgaa gttattatgc 70380
acaaagataa tttgttgacc gaggggtggcg catctaactt tttttttgta aaaaatgatt 70440

caatctgcac cccagctcta aatggaaata tacttcccgg cataaccaga gcattactaa 70500
ttaatgaatt aaaaaattat agtatagagg tcatcgaaga taacttcagg ctagaagatt 70560
tatcaaatgc ttcttgcgcg tggctcacia ggcgaacgaa agggctggcc cccatttctg 70620
aaattagtaa cctagaatct cacttggatt tataccatcc tcttttcaaa aagagtgaag 70680
aaatttttta taagaagttc cttagttaga agtcagttta ataactatat tatcaagctc 70740
gtaccaaaaa tctttatcac taagacctt atttgcaaga tcaagctcaa atacattttt 70800
ttggagctgt attaagtttt gtagtttatg atttttaata aaatgcatat attcagaagt 70860
tttatttttc caaacccag atttaattag actatctaca ggatttttac tttgtctagc 70920
attggctgca ctgtaataa ttttaccac tatccaaact agaagagggtg cataatgatc 70980
ttccgctgat tttattgaat ggataatttt taaagcttgc ttggtattat ttgaaataat 71040
tttatcttca agctcaaag gcataaatc tgcagactcc acagaagggtt ttctttcttg 71100
gccatcacca tctttataag tcagtcttag aagctttact tcattctggt gggcaactaa 71160
gtttccagag ttcatttcag atatattttg cactaaagag ggcctatctt tttcagaaat 71220
aaaagatagt tgatgcttaa gccaaatctt ttcttcaaaa gatttaagtt ttccacaatc 71280
aatgatgagg gctatctcat ccatctggtt taccactta gttgtgaaat taagcttttc 71340
attgcttgaa ttaataatta ttgcaatatt atccatatgg ttaatatatt caatttgaga 71400
tattttaatt atttggctcag gtatttttcc ttgatcatga ttaatctcaa ctattatttt 71460
tgagccaaat aaagatccag aagcattttc aataatagtc tgatctattt tatcaaacc 71520
atctttatta ataattgttt tttctttgaa ccctttattt gataaatggt tgagcaaaag 71580
atcttttagag ttatttttta agacaacctc agatccaaaa ataaagaaga tattttgtga 71640
ctgatctaaa tatttttgag ttgtcattgc ctcacatttc aaaagcatgc acctcaaaaa 71700
ttatttgatc aattaattcc ttctgcattt cagactcaag agatttaatc atttgatctt 71760
gtgcaaagtg attaagttca ttcgatttat atctcttgta gctaacaagc tttttactta 71820
tttgtttatt tccaatcatt tttatatgta tttctaaatt aatttctcct tcaagagctc 71880
ttagagaaga gcctccatag atatcatatt tatttaaaat atagtcttgt attagaatct 71940
gattttgacg attctcagct tctttttctg tattgaagag gattgctaac ttttgcctca 72000

aagaattagg gacactacta tcaaaccgaa aattaaatag atcttttttta tctagatcta 72060
tatttagttg attaaattgg cagccagtta aagatatcaa caagcaagtg tataaaaaga 72120
aaggctcttag ttccataacc tcatagtatc tatctgaaga atatatatca aagattaatt 72180
ttaaataaca aaatttatta ttttttcttt aatataaata gtttttttaa tagccgagtt 72240
gtttattggt gatgccacat tgtcgatagc cagtgcataa gcttctatat ctttttgctc 72300
aaggtttttg tctatcatta ccttacctct cacctttcca tttacttgaa ctattaaatt 72360
aaattcagaa acttctaata actcctctct aagaacaggc caagaagatt caatttcttc 72420
ttgtgcaaag tcaaagtaaa aattattcca taaatgttgt gagatatgag gcgcaatagg 72480
attgagagtt ttaagaataa ttattattgc ttcattaaga caatattgat ttgagatcga 72540
cgcattacct tctttaaaag aatcaggtat aaagttaact agctccatta tcgaagcaat 72600
ggctgtatta aatgaatatc ttgtttcaaa gtcatttgta acctttttaa gagtattatg 72660
cgattttctt cggagctcta cttcttggtt cgaaggatcc ttgggctcct ctaagtcaat 72720
aaattttcta ttgetgacta agttccatat ttttttcata aatcttgagg caccttctac 72780
agatgactct gaccattcta gactttgctc agggggagct gtaaacaatca tgtaaagcct 72840
tacagtgtca gcgccatact tttctatata agattgagga tctacagtat ttccttttga 72900
tttgacatt ttagcccat ctttaagaac catgccttgt gtaagaagct ttttaaaagg 72960
ctcattgcct tcaactaaat ccatactct aagcgcctta tgaaaaaatc ttgaatatag 73020
taagtgcagt atcgcatgct caatacctcc tatatataaa tcaacaggca gccaatattt 73080
tgaattttta tcaaacattt catctgcatt gtcagatgaa gtaaactctg cgtaatacca 73140
tgatgagtcc ataaatgtat caaaagtatc tgtctctctt tttagtcgat cagaaatatt 73200
ataaaaaatc tcattttgac ttaagggaat aggtgccgag ttctttttta gctctggaag 73260
cactataggc atatcctttt catcaataac tcttggttca ccattttcat aaacaaccgg 73320
aatagggcaa cccaatatc tttgtctgct cacaccccag tctcttaatc tgaactggat 73380
taactgctca ccaaggtttg cgtctgctaa atctttaata atttcaagag atgcctcctc 73440
agagtccatg ccatcatatt tatcagagtt tattaatttg cttttttgta caactggaag 73500
ttcattatta ttatcaaaac taataacttg aggaatctct agattatatt tagatgcaaa 73560

ttcgaaatct ctttgatcat gagcaggaac acccatcact actccagttc cataatcaag 73620
aagtacaaaa tttccaatcc atacgggtat cttcttcttt gtaattggat gaattacatg 73680
cattccacta ttaattccaa gtttttcagc ttttagccata tcagcttcag cagctttcac 73740
ttctttgcat ttgtttagaa agtcttttat cgattcatta ttttttgaca accctattga 73800
tatagagtga tttggtgata ttgctaaaaa agaaacacca aaaattgtat caggtcgtgt 73860
tgaaaatacc tttaaagaat catcggaatc ctcaatcatg tatttaattt ctgctccctg 73920
agatttccca atccaattcc tttgcattgt tttaacattt tcaggccaat ccacctcatc 73980
caaagatgtt aaaagttcct cagcgtagtc tgtaattttt ataaaaccatt gatcaatttc 74040
ttttatttca acttgcgccc cagacctcca tctttttcca tctataactt gtcattttgc 74100
taaaactgtc tcatcaactg gatcccagtt cactaatgat tttttcctat aaaccaaacc 74160
tttatcataa aactttttta atattaactg ctcccactta tagtattccg gttcacatgt 74220
tctaagctct tttgaccaat catacccaag gcccaaagat ataagttgct gtttcatatg 74280
ttcaatattt tgattcgtcc aatcttttgg gctaacttta ttagcaatcg ctgcgttttc 74340
agcaggaagc ccaaatgcat ccagcccat aggctgaaaa acattaaagt tattcattct 74400
tttatatcta gatatgacat caccgatcgt atagtttctt acatggccca tatgtaattt 74460
acctgatgga tatggaaaca ttgataggca ataaaatttt tctctaccat caggattagc 74520
tttaaatttg tcttctttta accattcatc ttgaatgggt tctctatatt gacttggggt 74580
atattcagga ttcatttttt taggaactca ttttctaagt aattaatagt atgtttattt 74640
tctttaaagg tatcatcatg cagaatcctt tgatgaagag ggtgatttgt tttaatgccc 74700
tctataaaaa attcatctaa agcactcagc attctcttaa tagccgagct tctagaattt 74760
gcctgagtaa tgatttttgc tagaagtga tcatagttag ggggaactct atatcctcca 74820
taaataatgt agtcatatct tatgccaaaa ccgccagggt tatgcatttt tgtaattggt 74880
cctggggatg gttgaaaatt atcaggatct tctgcattaa ttctgcattc tagtgagtgg 74940
ccatgaaaat taatatcatc ttgatttagc tcaattggca tttcaagagc aattcttagt 75000
tgtgctttta ctaaataaaa gccagttatc atttctgtta ccggatgttc aacctgaatt 75060
cttgatttca tttctataaa atagaattga ttatcttcgt ataaaaattc aatagtgcc 75120

accoccttcgt aattttatttc ttcacataaaa ttgacacaag ctatttagtgt tttattaaga 75180
gcttcttgat ttatattaag tgctggagct tcttcaatta ttttttgatg tctcctctgc 75240
atactgcaat cccttgtagc taaatgtatc gcctttcctt taccatcacc aactatttga 75300
acttcaatat gtctagggtt cccaataaat ttttctagat aaatggtttc attgccgaat 75360
ccatttttag ctctctgcat tgttatttct gcatgtccaa taagatcttc ctctctttca 75420
acaactctca ttctctctcc accgccacca gccgttgctt taatcataat tggatagccg 75480
atatctctag caattttttt aaactcatca ttatcagatg gaatctcatc tttataacct 75540
ggaacgattt gaattccaga tttttcagcc agtggttttg ctgttatttt gtcgcccatt 75600
ttttggattg tttctgaggt tggcccaata aatttaaate cacttttttc acacatttct 75660
gcaaaattat gatcttctgc aaggaacca tatccaggat atatagcatt agcacttgta 75720
agttctgcag ctgagagaat ggcaggtata tttaggtagc tttgagttgg agatgcaggc 75780
ccaatacaaa cagtttcatc agagaatctt aaatgtttta gatctttatc agcttctgag 75840
taaacagata ctgtcttgat accaagctct ttacaagctc ttatagctcg aagagcaatc 75900
tcaccacgat tggcaataag aactttataa ctcattttaa ttgacagtta taattttttg 75960
tccaaattca accggacttc catcttccac gtcaataact gaaatcttgc catcaaattc 76020
tgatttaatt tcattcatca ttttcatggc ttcaactata cagagcacat cacctacttt 76080
tatattgtct cctactttta caaatggatc tttttcgga cttgggcttc tataaaaagt 76140
tctactatt ggagaggtaa caacatcccc tatgacagtt tgcttagcct catccaattg 76200
atgagttggt ggtgagatag ctggctcatt tttgacaatg attggttttg aaacaaattc 76260
ttgattagag ccgttatctc ttgaaattct tactgactca tctccttgac ttacttctat 76320
ttcttttaaa tcagactctt gaagcatttc tataagtttt tttattttcc taatatccat 76380
ttattttacc ctgtatttta atatgccctt aagcatcgcc tcttcataac ccttagcacc 76440
aataccagtt atcacttctt cagctatatc agatagatat gagacatgtc taaattcctc 76500
tcttgatat atgtctgaaa ggtgaacctc atagaatggt atgttgacac caagaaatgc 76560
atcccttata gctatgctcg tatgagtata tgcagcagga tttattatta caaagtctat 76620
ttttcttca atagtgtat gtattgattc tattatttca tgctccgcat tactttgaaa 76680

agatattaaa ttgcattgat tttgatttgc taagactaat aaatcttggt gaatatcttc 76740
aagtgatttg ctgccatata cttccttttc tctagttccg agaagattga gggtggggcc 76800
atttataagc aatatattca tataatagat tttaacagaa atttacagat tttttaggat 76860
tttaaagata tttttataag agttgatttt tttgcatggg atagcaaaaa cccgcttcag 76920
agcatccctg atagtatatt aaaacctccg ataagttcaa tgaagaattt atttcaatgc 76980
ttataataaa ttcatctcta agaatctcag tctcgccaaa aaattcatcc ctatatagat 77040
ctttacttga ctcaagcgtc ttaaattggga gcactttatc tttaaatttg agctctatac 77100
tatccaagta catataataa ccatctctta tttgccaagt tatatatgct gcttgttttt 77160
caacattggg tgttaaagca aatacttcat tggcttctgg gaccctattt gagttttcaa 77220
aaagatttgc tgaattcaat tcacctgcaa tcacattgga tagtagtata attattaaac 77280
ttttgatcat gaagacttaa gaataaaccc ctaagtactt aaaataaaga ttttattaaa 77340
taacacggta gttaaaaatt tatgttccag actttcgcac taaaaatttt cttttatatg 77400
cctatttggc ttttaaaaat aattttttat agaaaaagaa cagtaataag agggcatcaa 77460
tttgatgctc aatctgctgc attattgtct ttgcttccaa agaaagattt atctgaatta 77520
ttagatgggt aaattgcaaa agctagaatt actcttgaag aagcaagaat tcaaaaataaa 77580
gtttctttaa caccgtctat acaagtcaga aaagtagatc atattttacc aaagcatgat 77640
ttaattctca gagaatacaa gccgcatcaa gaggatttaa aaaaagttat tctctatttt 77700
catggaggag gttatgtcct taactcagtc aatacacatg atgatatggg ttcatatatg 77760
tcagataaaa taggagtaaa gttttattct ctagactata gactatcgcc tgagagcaaa 77820
tatcctgatt ctcttgatga tgctctagat gctttttctt ggctcatcgg ccagggatat 77880
ggaccaggtg atatttcagt ttgtggagac agcgcaggcg ctcatctagc tgcttctttg 77940
tgccattacc tagctgaaca aaataaagat ttgcctagta gccagctact aatatatcca 78000
atgtgtgata cgtcctgttc atctgagtct tataatttat tatcttcagg atatcttcta 78060
actaaaaaaaa ctatgatttg gttttgggat aagcttagaa cttccgaagt aaacaataac 78120
gattcggcct ttaatctctt aaaatttaat tttgaaaaaa ctttaccgcc aacaattatc 78180
gttacagggt gctttgatcc tctatgcgat gatggagaaa aatatgcata tttattacat 78240

aaaggtaaac ataatgtgaa acaattacat tatccaacaa tgtttcatgg gtttgcacat 78300
atgactaaat tgaaagcagc gcagatagcc gttgaagatt ttttaaaaga atataagaaa 78360
atactatgag taaaatttta gaagtaagt acttaagcat caatttttca acaagggatg 78420
gattgtttta tgcagtggat aatataagtt ttgatataga aaaaaatcaa accttggcct 78480
tggttggtga gtcaggttct ggtaagtcgg taactgctat gtcaattctt cagctccttc 78540
aaaaaccaca agcatcatat tccaaagagt cttctattaa gtttaatggc gatgagataa 78600
taaagccaa gtatgaaaag ttactttcct tgagaggaaa tattatatct atgatatttc 78660
aagagccgat gacctcacta aacccttctc acagagtagg taatcagata actgaatcaa 78720
tactacttca ctcaaaaagc tcaaaaaaag atgcaataga tgaagcaaaa aaattaatgg 78780
cacttggtga gattgatgat gttgaaagac ggttctatgc ataccctcat gagctttctg 78840
gagggcagcg acaaagagtt atgattgcta tggcccttgt taataaacct gagctattga 78900
ttgctgacga gccacaacc gctcttgatg taactatcca agccagata ttagatctca 78960
tgtctaagct taaaaatgaa ctaggcatgt caatactttt tattactcat gatctaggcc 79020
tcgttcaaga attttctgac aatgtttgtg ttatgaagaa tggcaagata gttgaacaag 79080
gaaatactgt tgaggatattt aacaaccct ctcacgaata tacaaaaaaa ctttttagatg 79140
cagaacctca gcccaaatta gataaccctg taagtgatga gccataatt gagatcaatg 79200
acttaaatgt ttactactca ataccttcaa ctaatttttt taaaaagaat acttttcatg 79260
ctgttaaaaa tacttccttt agtatattata aaaatactac aattggcctg gttggggaat 79320
ctggatctgg aaagtcaact ttgggtaagg ctatagcaaa cttagtttct tataagggga 79380
atattaaatt tgagggaaga gatatcaact caaattccca aaaagaaaat aaagaattga 79440
aaaaaaatgt ccaaattggt tttcaagatc cttatgggtc attatcacca cgaatgacag 79500
tgggggagat agttggtgaa ggttttaggtg ttcactttta gcttacaaaa aaagaagag 79560
acgaaaggat agataagggt ctgtcagatg tcggtatcga aatagtagct aagaataaat 79620
atccgcatga gttttctgga ggccaaaggc agagaattgc aattgctaga tctttaafaa 79680
tgaatcctgc ttttatgatt cttgatgagc caacatcagc attagatagg tctattcaaa 79740
ttcaggtaat cgatttattg aaagagatac aaaatgaata tgggcttact tatcttttta 79800

taagtcacga tttaaagggt attagatcga tgtcagactt tatttttggt atgaaaaatg 79860
gagaaatcgt agagtcagga ccttctcaca aggtctttga aagcccagag caagactata 79920
ctaaaaaatt actatcagct gctttaaagt atgcatctga ataattaaat atatymcmtm 79980
tggcaaatag aaartattca aaagagctcg ttgacgggtcc taatcaagct gcttctagat 80040
caatgcttag aggagtaggt ttcacatctg aggmtttcac aaaaccattt gttgggattg 80100
cttccacagg agcaaaagta accccatgca atatgcacat aatgcactt tcagagatcg 80160
ttgagaaatc agttgatagt tcaggaggaa aggggtgttct ttttaatact attactgttt 80220
ccgatggaat ttctatgggt acacagggtg tgaaatatcc tcttggttct cgagaggtaa 80280
ttgcagattc aatagaaact gttgtgggat gtcttggtta tgatggagtt attgctgtcg 80340
gtggttgatga taaaaatatg cctggatgca ttattggaat ggcaagatta aatagaccat 80400
caatatttat atatggtggt tctatcaaac ctagtaaaga aaataccgac tatgttactg 80460
tttgtgagaa aactggagag tactcaaaag gcgatcttaa agaacttgaa ttaattcatg 80520
tagaaaaaat ttccgtaaaa gggcctggat cttgtggggg aatgtatacg gcaaatacta 80580
tggcttctgc gattgaagct ctaggcacga gtcttctctg aagcagcagt caagatgcaa 80640
tttcacacga caaagaagat gattgtttta aggctggcga agcgataatg aatttattag 80700
aaaaagatat taagccttca gatattatga ctaagaatgc ttttgagaat gctataacaa 80760
tggttaattgc tctaggaggt tcaactaatg cagttctgca tttattggcc atggcgcatt 80820
caataggggt tgatttagag ctagatgact ttacaagaat aggaaaaaaa acaccggtta 80880
tggcagatct taagcccttt ggttctcatt atatgtctga actcaatgct aatggcggta 80940
ttcagccact aatgaaaact ttgcttgaga agggattact acacggcaat tgtcttaccg 81000
ttaccggtca gacgcttgct gaaaatcttt ctggaataaa accttacgag cctgatcaag 81060
agataattaa atcatttgat aatccaatta aatcaaatag ccatcttaga attctgtatg 81120
gcaacttagc gaaagatggt gcagttgcaa aaattacggg taaagaagga acttcctttg 81180
aaggaagtgc tcgtgtattt gattcagaag aagaaggggt taaagcaatc ctatctaatt 81240
ctataaaagc tggagatggt gttgtaatta gatatgaagg gccaaaagga ggtccaggca 81300
tgagagaaat gctaaaacca acatctgcca taatgggtca aggtcttggc gatcaggtag 81360

cttttataac agacggtcgt ttttcaggag gcactcatgg atttggtggt gggcacatta 81420
ctccagaggc tgcagatggc ggcttaattg caataattaa agatggcgac tctatattaa 81480
ttgatgcaga ttctgataag ttaattctta atatttctga ggatgaaatt tcaaatagac 81540
taagtagatg ggtaaaccce aaaacgcctc ccaaaaaagg agtcttagca aaatttgcaa 81600
aaagtgttaa atcagctagt cttggagcgg taacagatta aatatgtatt ttaaaagaaa 81660
atttccaaat agtagactaa gaagaatgcg tctcaattca aacctcagag acttgcttgc 81720
tgaagttagg ttgtctacaa atgatcta atcagccatta tttataaaag aaggcttgag 81780
tggaagagag gctattgaaa gcatgcccac tattaataga tatggacaag attcaatttt 81840
ttcagaaata gaagagctac tagagcatga tataaatagc attgccctat ttccagttat 81900
taatgattcc aagaaaaaca gcactggaga tgaggctatc aatgcactca atttgatgtg 81960
tgaaactatc agtaacataa agaaacgatt ccctgaaata atcttaatat ctgatgttgc 82020
tctagaccca tataccgatc atgggcatga cggtatctta aaaaatgatt atgttgataa 82080
tgatgagact ttagctgttt taaggaagca atcactaaca ttagctcagg ctgggacaga 82140
cataattgca ccatcagata tgatggatgg aagaataggc tctataagag aggcattgga 82200
cgaaatggat tataaaaata caatcctact atcatatgca gcaaaatata attcaaagtt 82260
ttatgggtcca tttagagatg ctgttaattc agcttcaagt ctaggcaagt cttcaaaatc 82320
cagttatcaa atgtcaccaa aaaatataaa tgaagctcta catgaagttg ctatggatat 82380
taatgaaggt gcagacatag tcatgggttaa gccaggtatg ccttatctag atataatttc 82440
aaaagtaaaa gaaaccttca aagtacctac ctttgcatac caggttagtg gtgaatatag 82500
catgcttaaa ctggcgattg ataaaggatg gcttgaaagc gatgttatgt tagaatcatt 82560
aataagtttt aaaagagcgg gagctgatgc aattctaagc tacgcagcta aagaaatttc 82620
caaggagata actaacaat gagcaatgtg atagaaattc gtgatgaaga aagctttaat 82680
agcgacgtct taaattcaga aaaacctgta ttgggttgatt tttgggctga gtgggtgtgga 82740
ccttgtaaac agcttgcacc aactgttgaa acagttgcag cagaaaaatc agaaacatta 82800
aaggtttgca aaatggatgt tgattcaaat agagagattg ctgctaaata tggaataagg 82860
tcaatccctt cattaattat atttaaaaac ggagagcctg caggagttga agtaggtgct 82920

ctaaccaaac aacaattaga ggactttata agtacagtag tttaactttg caaagacttc 82980
tttgcatatt gcaagaaata ggattatcat ttcgacttct aaggctgaaa gccactcaaa 83040
acaaaaaccc acctttttcaa tcataaaaaac aacaactaga acggaaataa ctaaataaac 83100
cttactgaaa ttaaaataaa accaataaat gaacttgtag atatagctac tgagcttggg 83160
cttgaggatg ttggaaggct gaaaaagcaa gagataatat ttagaatatt taagcataag 83220
gcttctgaag gtgttgatat ctatgggtggc ggagttcttg agatttttaa tgatgggtttt 83280
gggttttttgc gatccccaga aggctcttat tgcgctggcg aagatgatat ctatgtttca 83340
ccaagccaaa taagaaagtt tagcctcagg aaggagact cagttgctgg gaagataagg 83400
accctaaag ataaagagcg ctattttgca ttaatccaag ttgatactat taatgggtgaa 83460
gagccaagaa agactaaaaa caagattctt ttgaaaatt taactcctct ttttcccaat 83520
gaaagactaa tccttgaaca aggaacagggt tctaataag atctttcatc tcgaataatt 83580
gatttgattg ctccaatagg aaaaggctcag cgtggactta tcgtttctcc acccaaggct 83640
ggtaaaacct taatgcttca aagcatagct cattctatta aaagcaataa tccagaagta 83700
gagcttatag ttcttttgat tgatgaaaga cctgaagagg taacagagat gtcaagaact 83760
gtaaaaggag aggttgtagc tagtactttt gacgagccac ctactcgaca cgttcaagta 83820
gcaaatatgg ttattgaaaa agcaaaaaga cttgtagaac ataagaaaga tgtagttatc 83880
ttattagatt ctattactcg tcttggaaga gcatataact cagttcagcc tgcacagga 83940
aagatattga gtggtggagt tgactccaat gctcttgaaa ggccaaaaag gttttttggt 84000
gctgctagaa atcttgaaga ggggtggaagt ctactattc ttgctactgc tctagttgaa 84060
acaggctcaa agatggatga agttatttat gaggaattca aaggtaggg taatatggag 84120
attcaccttg aaagaaaaat agccgaaaaa agaatatatc ctgctattaa tattagaaga 84180
tcgggaacaa gaagagaaga tttacttact gctgaagatg aattacaaag aatgtgggtc 84240
ttaaggaaaa ttttagacga tatggaagat gctcagtc aaatagataga 84300
ttgaaatctc ataaaacaaa cgatgagttc tttacttcaa tgaaaggggg taatggcaag 84360
aagactagat aaagtttttt gccatatcaa tcatcatctt atcagttggg ctttgcgata 84420
cgtgaatttc taattcttta aactcatttt gacatctatc ctttatattt tctgaagcaa 84480

caataaatat tttttttctt aaaactgcat cgtccagatt tgttattaag aatttaagag 84540
tactaaaatt atagattaaà aaaatttcat tattatctgt aacttttggt atttgctcca 84600
gcaaatagat tacttcatag caaactatct catctagact agccttaagt ttttggtgaa 84660
gaaaaccatt tgaattttcg ccgcaaaata aaagactctt tcctagaaaa tttttctcta 84720
ttaacttaag aattccttca gatgaatggc tttgtggaaa atgagatttt atgccacttt 84780
caagaagttt attggaagtt gcggggcccaa cagacaaaaa attgtgtgga agatcatcta 84840
aatcgaaaaa agatttttaa atatcaagtc cgtacgaagc agcagcttgg ctagtaaaaa 84900
ttaagtttga atatgaatga atatttttaa ttttatcgat agcgattttt gaagggtcga 84960
ttgaattaat tttagaaaga taaatatttt taagagctat ttcttgagct tcacataaag 85020
atattaggtt tcctgataaa tttttgggcc tcgtattaat tatcattttt taaaattaaa 85080
gatttcgctc cttccaaaat aaactcatca gcaataaca taatatctat tgcaatatct 85140
tgaaaggaag atatcttttc tttataaatt ttttcgttgc cttcataaga taatattttc 85200
cctgatattt ttatttggcc atttttatct tcacataaaa tagctattgg agacaggcaa 85260
gtaccttcca tggcagcaac aaaagatcgc tctgcaactgg caagaattaa ttctttaggg 85320
tcaccaatat tttctaaaag ctcaataata tcttttttat ctgacagaca ttcaattgct 85380
atatatccct gagatgcaga aggtaacatt tcttcaattg agaattcata tgaattttgt 85440
tttaaaccac gcccttttat agcagcctta gcaactacca gcccatcaaa cagaccatca 85500
ttgagtttct ttattctagt agctatgttg cctcttactg gaatagtttt tatatcaggt 85560
cttaagttat taatttgtgc ttttcttctt ggaccagaag taccaatagt tgaatttcta 85620
gcaaattcag aaaatgattt cccatcctta gatagaagca ggtcttctgg agattcccta 85680
ctaaaaacac ttattatttt aaactttgga tcaagctttg ctggaacatc ttttaacta 85740
tgcactgcaa tatctgcttc atcagcctct agagaggact caagtgtaga aatgaatagg 85800
ccctttccac caatttcatg aagaggttta tctgtttggc ctccttctga tgtcatagga 85860
acaagctcaa ccttaatat attgatctta gctaataact catctgcaac aaattttgcc 85920
tgatacattg ctagttctga ttgccttgta gctattctta ttttcatttt gtttccaata 85980
aaagagcctt cacatcgcca acactctttt ctttcaagat agtgaattcc tcaggaattt 86040

ccaaggtagt aaatttacta tactcaaggt agattttaca agatggcttt atttcattct 86100
ttcgaattat agattttaga acttttaatt catattcttc gccaaaagga ggatcaagaa 86160
gaattaaatc aaaacgagat aagtcatttt ttttaatcca actaaatgca tccttaaaaa 86220
aaactttaga tttatccttg atgccaaagaa gttgaatatt ttttgctaac acagaatagt 86280
tttttttatt aagttcaacg aacacaactt tttttgactg tctggatatt gcttcaatac 86340
caagcgcacc cgttccggca aacaaatcaa gacatattaa attttcaatc tcaaattgaa 86400
gccaaattaa aattattttct ttttaatttat ttgaagttgg tcttaaggaa tccttaaatt 86460
cgaaaggtat ttttttacct tttaaataac ctccagtgat tctgatattg tttttcattt 86520
tttttgcaaa atgggtatgc ctatttatta aataaatata attamaaata catttaaaact 86580
aagttataat tcatccatgm gtccacaag ataatttttag aaaagcaatg aggagctata 86640
tttactctgt cagtgtgatg tcaaagtgtg acgagaataa aaagtttagt gctataacag 86700
tttcttcagt tacctctgtt tctttagatc ctccagtgtt gctcgtttgt atcaataaat 86760
cagctggaat tcacaactca ataaaagaag ggtcctcttt ttgcataaac cttttaaata 86820
aaaatcaaga agatattttct aatctatgca gttcatttaa gaccgaaggt gatagattta 86880
atagtgggtga ctgggattta agcggcactc cgttttttaa aagtgccag gctaataatt 86940
tttgtactgt tgatcaatta atttcatacc acacacatac tattgtgatt ggtcatgtca 87000
caaactctct tagcgatgag aaaattaata cactgacata tgttgatggg agctatgaat 87060
aaattttcaa aaaatgtatt ttttatttta ataatactca actctttttt tcttgccctc 87120
aacatttttg ctagccagga agagtgtgaa gaaaagccaa gtgtttttat tatctctcct 87180
caagacgggt ttatctcaga atctaataat gtaaaagtct tatttggaac aaaaaatatt 87240
gaaataaatc cagctggcaa aggtgagatt gcaaaaaata aatgttttgc aagcggggcat 87300
caccatcttt tagttaatat cgaagcattg ccagagagct ttattccttt tgacaagggg 87360
tatttacatt ttggaggagg tcaaactgaa acaattcttg atcttgatcc tggaacctat 87420
tctctccaac taattcttgg atcttatgtg cataattcaa aaatgcaggt aaataacttt 87480
aaaggtcaag gacccttttt atcagaaaaa ataacaatta cagttaatta gagattagac 87540
cagttataac tttatctaga tgttcatcta agttattagc aattatattt gcactagaaa 87600

taggaccaga attatcgtca taatcatcac caactacatt cacctttctt attatcccaa 87660
caagaccatg caacatagac caaagagtta tacatttaaa ggcaatcact tcctcaggct 87720
cgtctgctag attagcaaaa ctttttctca tattatcgta tgttccattt gcagatttaa 87780
gaagatctgg gtaatcggca aagttaccaa cagctgttcc aaacatcaaa tcatatgtat 87840
gtgcattttt taaaccaaac cctatatatt tgcttgcat agtaacaagt tttttctttg 87900
taattttttt tggattctca aaaaaaaca gctcattaag ttttttaaag ccaacggttg 87960
caacagcggc atatacacat tcttttgttt caaaatgcct ataaggagca gtttgagaaa 88020
caccactttc ctttgcaagg gatctaatac ttagtttagt gtaaccatct ctatcacaaa 88080
gcctgcatgc gcattctata agttcttctt ttaagtttcc gtgatgatag ttattcataa 88140
tttaatttta atgttgacac tgctaacata tatcatgtta ttatgtatac accgcataca 88200
ttaagttgat tacatattaa agtcaaatac agatatatga acataacaaa aataaataac 88260
accgtcatag ctcttttatt aggcaatgcg tttctatcaa atcttgaggc tttagaagtt 88320
cttgaggtta aaatgcttga tgaatatgct gtaactagag aatttcttgg aaagctcatt 88380
ccatcagacc agtctaagct agcatttgaa atacctggaa agataaactc tattaatggt 88440
gatatcggag atgaggtcat cttaggggat gaacttgcct cattggatga tagagaagct 88500
ttagcacaat taaatcaatc aaaggctaaa tttgatttag ctgaacaagt actagcaaga 88560
tatataaatc tcagagcaga tgggcatatt tctattcaag atcttgataa ggctgaatct 88620
gatcaaata tagctaagtc gcagtatgat ttttatagag ttaaatttga gcaaactaag 88680
ttactagctc cttttaatgg agttattcaa aatagatttc tagatacagg atcagtaatc 88740
aatgcaggtg tccaagtttt agaaatttta ggctctagca atgttgaagc aagaatttct 88800
attccaatga actatatgaa caaggttaag attggagatg agtatgaatt tgatatcaga 88860
ggaatatcta caaaagctac gttagagaga ctggctccca tgtcaaccgg aggctccaat 88920
aataggttag caatttttag atttgatacc ttttttaatc caggatcaat agtaaagctc 88980
aaactaagca tcaactgagaa agcaaaaggt acatgggttc caattaagtc actgtccag 89040
tccgaacaag gtatttgggc catatatacc attaacgagc aacaagtagt tgttagagat 89100
cttgttgatg ttattttatt tgaagacgaa tatgcttttg tcagcggaac acttaataat 89160

ggtgatttag taatttttagg cggagctcaa aaaattattg aaggaaaaat aataaaataa 89220
aatgaatggt attaattttc taatagaaaa gcctaggata ttattttctaa ctttagcatt 89280
tatattactt tctggaattt cttctgggct ttcagttcca attcaagaaa accctgaact 89340
ggctgagaga tggggagggtg ttctgtattt tcttcccggg gcctcccgag aaagaattga 89400
aacagagata gtaaattgatc tagaaatcaa acttagagaa gttgaagaaa tcgatgagct 89460
tgaatcaatt attactcaag gtttttcaac aattgtagtt gaattaaatc aaagtgtacc 89520
tcctatactt attgaagaga cttgggtccaa gggtcaagac aagctcaatc aaatagttat 89580
tcctcaaggt gcagaaatat ttcttgatag aagcagtggt ccgcctatca ctgttcagta 89640
tgctgtaacc tggaacggca gtggagatgc tccactaata atgatgtcca gactagcaag 89700
ccagctaaaa agaaaattaa gctcaatagg ctcatctcat caaactgcaa tttttggtga 89760
aacagatgaa gaaattttta ttgaactaga ttcacaaag ctatcttcgc ttggattatc 89820
atttcaagat atcgcaagtgc ctattcaagc cctagatgca aaaaaaccta ttggtgtatc 89880
ctcaaacaac aattctgagc ttttatatag actcaaagat aatatacaga gcattcaaaa 89940
actctcagaa atacctatca aggttattaa taaatcagag atcatacagc tagatgatgt 90000
ggcatttatt tcaaaaatcc cggtttctcc tattgaagac atattcttgt ttaatggaaa 90060
tgtagttatt tctgttgctg gaaccggatc attttctcaa agagtccatg attatgtaga 90120
acgcgcaaca attgtttag atgagatgag agaaactctg ccgactgaga tcaactataga 90180
tttagtttat gacgaatctg cttacacaac taaaaaattt aatgagcttg taaaaagttt 90240
ttcattagca atattttttg ttttagcttt aagtcttttt tttcttgga ttagatcagc 90300
aataattggt actcttatcc tgccattttc tatttgctg gttatgattg gttgtaggtt 90360
tataggctta cccttgcata tgacatctat cactggaatt attattgcac taggattgct 90420
tatagataat gggattattg ttgttgaaga ttataagaat agaagagcat ctggattaaa 90480
tatcaatgat tcaatttcac aaggactaaa aaacttatgg gctccattag ctgctgcaac 90540
agcaacaacc gtcttctctt ttcttcctat tgttactgga gaagggtcga gtattgaatt 90600
cgtaggcggc atggcaatga cagtcattat gtctataaca tcttcattag ttttggcggt 90660
attaatgggt ccagttctga tgagttatat ggaaaaaatt ccgttcttta aggatgtgga 90720

tattagcaag gaaggggtata gaaatgaaaa aatccttaaat aaatataggg ccttttttaaa 90780
ctgggcgttc ttagttccta gaagagcaat catgatatcg cttgcattgc ctggttctagg 90840
attcttttctt ttttaattctt tacctaaaga tttcttttctt gctcaagata gagatatgtt 90900
tagagttaat atagaactgc cttctaacgc ctcatcactt actacaatgc agagagttaa 90960
ggaaattaga gaagatatct tagatagtga ttttaatttca atagaaaaag attattcgtt 91020
tatcggcaga atgatgccta gagttttgat gaatgttggt ggtggagaag aaaaacaagg 91080
atccaataat attgcgcagt ctgtattttt tgctactgat tattatgaaa tgattgaaaa 91140
ccttccagat ttatcaagaa gactgggttaa aaataaccct gacattatag ttaatatattga 91200
tagttttctcg tctggccccc cggttttttt agatgttagt tatgtaattt ttggagatga 91260
tccagattta ctaaaatcac ttggtgagga gctagagcta attattaaca attctcctga 91320
tgtgagtctt acgaaatctg caacttcaaa ctcaataacc aatgttgagt ttgaacttaa 91380
cagctcaaat atttcactgt ctggtcaaaa tgccaattat cttgtaaatg aaatgtttac 91440
tgcaaacaat ggaatatattg ttggcactat gttggattca aacaaagaaa taccagtcag 91500
gctgaaaggg ctgtctaata aaaacaatat tacgggaaat actagtttta taacaatgcc 91560
ctctcaaggt ggttttgagt attttgatag tttcggaaaa agctcactaa caaacaatc 91620
gtcaacaatt actaggcttg atggccaaag aacaaatgat gttgagggct ggatttggac 91680
aggtacgctc ccgtctgcta ctgaaaaatc tattaaaaaa gatgttaaag attttgaatt 91740
aagattgcca ataggctatt cattaaaaca acttggcgag gctgaaagca ggggccaatc 91800
tcaagcctca ttatactctt cagcttttat gtatttcatt cttataatag taggcttggt 91860
tatggcgctt aattctttca gagaggctgg tctaattttg tctgttgcat tcttatcaat 91920
tggaactatcc tttcttggtt tatttatagg ccagcaaaat tatggattta ttggaactat 91980
aagtgaata gggtttaattg gcttatcaat aaatgattca attattgtct tatctcacat 92040
aaaagaagag gctgagaaga aatcactaac caaagctgag cttgttgaag ttgttatcag 92100
gtctacacgt catataatca ctacctctt gacaacactt ggtggttttt ttcacttat 92160
ttttgcaagt gtattcttca aaccgcttgc ttgggcaatg agcattggag tattaggcgc 92220
gactattaca gccttattat atattcctgc aatgtttatg ataatgagaa aggttaagta 92280

ctagaacaac tttccgagca ataatttctt cctagctttt ttacaataag tgattcatga 92340
gtataggtct cgcaagaatc acattttaatc atagagttat caatcttttt tttatgaggt 92400
cctgatatga gagatctgaa aaagaatacc gcagttatta tcaatataaa tatgaccaat 92460
ggaatcaaaa gtaatatcga ttttaataag aacatttttt agtcactaga tgctggctta 92520
ctattttctta taaaattcca agaccagcca gacttatctt cagatttttt tgattgctca 92580
tcttggtagt taatagaaat tattttttta gtatcttcaa gcaattcaat atatcccaaa 92640
gattcatatg aagcctctaa gatcttaaga gctctataat tttcactaga gttaggaata 92700
ttttcaatta cataatttgc tcttcttatt gctgctatat gcgcatcaac actaacgtaa 92760
taatcagctg ccgcaagctc atttcttgca atcatatttc ttaaatagat atttctttgt 92820
ttagcatatg tagaatattg gctatcagga aatcttgta agaattcagt tagttctgaa 92880
aatgattcct tagctcctga gatatctcta tttgatagat cagtatctgt cattcttaca 92940
ataaagctat tatctctcgt atagctagaa aggcctttca taaaatatgc ataatacaata 93000
tttggtgccc taggatgaag tcttataaat ttttctgctg cagcatgcgc agcctcagtt 93060
tcagcattca taaattgagc ataaataagc tctacttgcg cctgttcagc atatttgcca 93120
aaaggatatc ttgattcaat tgcttcta ataatcaatag caccaaaaata attttttcca 93180
gccatccttc tttgggcttg atcgtaataa attttttcag gctgttctat ttctgggcca 93240
tcagaattac aactaaccaa taacagagtt acaattggca atactataaa tagttttaaa 93300
attagtttat ctttcattat ttgcacattc tacctgcatt atcgtaattt aagcttgaaa 93360
acatctatct agattaattt ttttataata agttcatatg attgttaaaa atgttccaaa 93420
tgatctatct tcaatgaggc tagacaaagc tacagcagaa atgtttacag attattccag 93480
aactcagata aaaaaatgga tagaagaagg cagagttctt ataaatggag aggtatccca 93540
gccaagagat aaagtttatg agaatgatca gattgaatta agccctaaag aagaacaaaa 93600
agtatcatgg gaagctcaag atatagattt tgaaattcac tttgaaaatg aagattttat 93660
tataattaat aaacctgctg gtttaataat gcatccaggt tctgggtgct atgatggaac 93720
tctcgcaaat gggctcattt ataaatttcc agaattgatc aatattccaa ggtcgggaat 93780
tgttcatcga ttagacaaag atacttctgg cattctgctc gtagcaagaa atgagtctt 93840

taggaacttt tttattaatg aaatgcagga gaggagagtt gtaaaaaaat atacgtctat 93900
tggtattggt tctacactag gaagcttttc tatagaagag ccaattggaa gagataaaaa 93960
taatagaacc aaaatggcaa ttcgagaaga tggcaaagat gctttaacat ttgtaaagct 94020
taaagaaaat attggaaact actctgtgtt ggatataaga atagagacag gaagaactca 94080
tcagattagg gtccatctat catcaaaaaa actaccaata attggagata aaacctatga 94140
ccaagcagg tctattgcaa gagatacccc tgaagagcta attaataatta tccgagggtt 94200
tccaaggcag gcattacatg caacacacct ctcatccaat gaccaaaaaa caaataatat 94260
tttttctttt gatattccca ctccaaatga tatggaggaa ctacttctag aattaagaaa 94320
attgatctaa tagtaactaa aaacttggtt ttgtattaat aaaataatat aaaccttatt 94380
cctaagagtt ttttggttaag aaattgaaat tatctggcgc agacatgcta atgcaagcac 94440
ttcatgatga aggtgttgag ctaatctttg gctaccagg tggagccgcg cttcatatct 94500
atgatgcaat ttttagacaa gataaaatag atcatatttt agtaaggcat gagcaagggtg 94560
caacccatgc agcagatgga tattcaaggg cgacaggtaa gccaggagtt gtcttagtca 94620
cttctggacc tggtgcaaca aatgctatta caggaatcgc gactgcattt atggattcca 94680
taccaatggt agttatttca gggcagggtg ctagccattt aatagggtact gatgcttttc 94740
aagaaactga tatgattggt gtttcaagac caattgttaa gcatagctat acagttttta 94800
atgctgaaga aatacctaag ataattaaag aagcttttta tgtcgcaact tcaggcagac 94860
ctggacctgt tggtatagat atcccaaaag acatgacagc tccggataat ctttttgatt 94920
actcgtatcc tgaagaagcc aagataagat catacaatcc tcogattgag ccagaaaaaa 94980
atcaaataga tagagcagtc gaagctatat tgatatcaaa aaaaccagtt atatatgctg 95040
gtggtggggc aattgctagt aatgccgaaa aagaattact tgaacttaat gaaattattg 95100
atgctcctgt tacaataact ttaatgggat tgggtattta ccctgctagt catcatagat 95160
ttcttgggat gttagggatg catggaacat atcaggcaaa tatggcaatg cataatgcag 95220
acttaataat tgctattggc gccagatttg atgacaggat taccaataaa ccatcaaagt 95280
ttgcacctaa tgccaaagt gttcatctag atgttgatca ctcatctgta tcaaaaatta 95340
tagaagcaaa tgtagctgtt tttgggcaag taaaaaatc cttaaaatta ataaaagaaa 95400

ctcttgaaaa aaaattagac tcttacgatt ctttcgctct tcagccttgg cacgatcaga 95460
taaaagaatg gaaatcacta catgggtttaa attatgagct ttataaagat gaatctgatg 95520
atcatcccat tttaccccag gctgtagtcc agcatgtcca tgagattaca aatggggaag 95580
catatgtgac ttccgatggt ggtcagcatc aaatgtttgc tgctcaatat tatcattttg 95640
ataagcctag aagatggatc aattctggtg gtctaggaac tatggggtttt ggtttgccag 95700
cagcaatggg tgtaaaactc gcttttccaa aagatgaggt tgtttgcatc actggtgagg 95760
gtagtatcca aatgtgcac ccaagaattgt ctacatgtct tcaatataat ctcccaataa 95820
aaataattaa tatcaataac gaagctcttg gtatgggttaa acaatggcag gatatgaatt 95880
atggaggaag gcaactctgaa agtacctatc aaaactcgtt accggatttt ataaaactgg 95940
ctgaatcata tggatcatata ggaattaaaa ttacaaaaaa ttctgattta agtgaaggct 96000
taaaaaaagc ttttgaaatg aaagataaac ttgtctttgt tgatatttat gtagatcctt 96060
cagagcatgt ttatccaatg caggttgcaa atggcagcct agaaaatatg tggctatcaa 96120
aggatgaaca aacatgatta aaagaaaact aattttaatt atggaaaata aaccaggagc 96180
tctagtaaga gtagttggac tgtttcatca aagaggctac aacattgaaa cccttcatgt 96240
agatactggt aaagactttt ctacttacia atcgatattg aaaaaaaact tgaaaccaa 96300
tttgaggata atcaaataat tagactgacc atagaaacia tggtttcaga tgaccttatg 96360
aggcagattt tgagacagct caataaatta atagatgtta tagctgtaag caatgaagag 96420
acaacctatt taaaaggagt attattagat gaaaatttat tatgacgaag atgcaaacat 96480
agaaattatt aaagggatga atgtctcaat aattggctat ggatctcaag ggaatgctca 96540
tgcaaataat cttcatgaat ctggtgtaag tgttactgta ggtttaagag aagggtcttc 96600
ttcatgggca aaggcagaag aagcaggctt aaaagtcca acagttgctg attcggtaat 96660
ccaagcagat ttggttatga ttttggcacc agatgaattc caaaaaata tatatgaaac 96720
cgaaatcaag ccaaacttaa aaaccagtgc aattcttgca tttgcacatg gctttaatat 96780
tcattttgaa aaaatagtct ctgaagcaac taacagcgta attatgattg caccaaaaagg 96840
tccaggccat actggttagaa gtacttatac caatgggtgga ggcgttccat ctctcatagc 96900
tatatatgaa gatgctttta gtgatgaaga ttattcagca aaagatgtag ctctatctta 96960

tgcaaaagca aatggcggca cgagggctgg tgttcttgaa acatctttta aagaagaaac 97020
agaaacagat ttatttggcg aacaagcagt tttatgtggc gggcttaccg ctttaattaa 97080
agctgggttt gaaactctag ttgaggcagg ctacagtga gagatggcat attttgaatg 97140
ccttcatgaa acaaaaactaa tcacagactt aattcaagaa ggtggcattg ctaatatgca 97200
ttactcaata tcaaatactg ctgagtatgg tgattatgtg agtggacca aagtaattac 97260
tagcgatacc aaaaaagcaa tgaaaggaat actagaaaat atacagtcag gaaaattcgc 97320
agatgacttc ttgaatgact gtcgacaaag caatgatggc actggcggac ctgttatgaa 97380
aagcaataga gaagctacaa aaattcatcc aatagagtcg gtaggggctg agctaaggtc 97440
taaaatgaag ttcctaaatt cacaaaaatt ggtggataaa gaaattaatt aaaattaatt 97500
aaaaaaaaagg tatcttcttc ggaataactt cgttaagata cgcgtccgca caaagaggcg 97560
gttggtcttt aaaaatattt gggttactcg gtgggtgttc aaaatacgag aaaaaataat 97620
ttagattttt tataaaaatc aacaaaacat gatattaatt gaagagtttg atcatggctc 97680
agattgaacg ctggcggtag gcttaacaca tgcaagtcgt gcgagaaagt atcttcggat 97740
atgagtagag cggcgggacgg gtgagtaacg cgtaggaatc tacctagtag aaggggatag 97800
cccggggaaa ccgggattaa taccgtatac ctcttcggg agaaagaagg cctctctttg 97860
aagctttcgc tactagatga gcctgcgtaa gattagcttg ttggtgaggt aaaggctcac 97920
caaggcgacg atcttttagct ggtctgagag gacgatcagc cacattggga ctgagacacg 97980
gccagactc ctacgggagg cagcagtggt gaatattgga caatgggcgc aagcctgatc 98040
cagccatacc gcgtgtgtga agaaggcctt cgggttgtaa agcactttaa gcagggagaa 98100
aaagttataa gttaatacct tataaccctg atgttacctg cagaataagc accggctaata 98160
tccgtgccag cagccgcggt aatacggaag gtgcaagcgt taatcggaat tactgggcgt 98220
aaagcgcgcg taggtgggtt gttaagttgg atgtgaaagc cctgggctca acctaggaac 98280
tgcatccaaa actaactcac tagagtacga tagagggagg tagaattcat agtgtagcgg 98340
tggaatgcgt agatattatg aagaatacca gtggcgaagg cggcctcctg gatctgtact 98400
gacactgagg tgcgaaagcg tgggtagcga acaggattag ataccctggg agtccacgcc 98460
gtaaacgatg acaactagct gttgggagac tatgtctttc agtggcgcag ctaacgcctt 98520

aagttgtccg cctggggagt acggccgcaa ggctaaaact caaatgaatt gacggggacc 98580
cgcacaagcg gtggagcatg tggtttaatt cgatgcaacg cgaaaaacct tacctactct 98640
tgacatactt ggaggctctt gtaatgagag tgtgcctttt ggaaccaaga tacagggtgct 98700
gcatggctgt cgtcagctcg tgtcgtgaga tgttccgtta agtcggataa cgagcgcaac 98760
ccttaccctt atttgccagc gattcggctcg ggaactataa ggggactgcc ggtgataaac 98820
cggaggaagg tgaggacgac gtcaagtcac catggccctt acgagtaggg ctacacacgt 98880
gctacaatgg gagatacaga cggacgctaa gccgcgaggt ggtgctaata ctaaaaagtc 98940
tttcgtagtc cggattggag tctgcaactc gactccatga agtcggaata gctagtaata 99000
gcggatcagc atgccgcggt gaatacgttc tcgggtcttg tacacaccgc ccgtcacacc 99060
atggaagtgg attgcaccag aagtagatag tctaacctta gggagggcgt ttaccacggt 99120
gtgcttcacg actgggggtga agtcgtaaca aggtagccgt aggggaacct gtggctggat 99180
cacctcctta acgataaatc gcgtttttaa cggccacacg agtaatcaaa tattaaaaaa 99240
aagaacattt agatatgtaa aatcattggt atgtaatttt ctagtgtata catttatgta 99300
tacataagat cactgcaatt aaaaagtaac atatgcattt atgtgtatgt taaaaaagta 99360
attaatatat ttatttaagt tactctcaaa aatgaagata aaacttcaaa aaaaatatgt 99420
aacctttttt aaggttatat gatcaagtaa aggaagagca caaggcggat gccttggcag 99480
cataaggcga tgaaggacgt aataacctgc gataagcctc ggggagctgg taaataagct 99540
tcgatccgag gatttccgaa tgggaaaacc caatacacat aagtgtatta tcttatactg 99600
aatacatagg tataagaggc aaacctaggg aactgaaaca tctaagtacc tagaggaaaa 99660
gaaatcaaca gagattccgg tagtagcggc gagcgaaacc ggaccagccc ttaagcttat 99720
tttagtccag caaaatatte tggaaagttt agccatagta ggtgatagcc ctgtatgcga 99780
aagactaatt taagtgaat cgagtaggtc gggacacgag aaatcttgac tgaacatggg 99840
gggaccatcc tccaaggcta aatactctat gctgaccgat agtgaaccag taccgtgagg 99900
gaaaggcgaa aagaaccccg gcgaggggag tgaaatagaa cctgaaacct tgtgcttaca 99960
agcagtcgga gcagacttgt tctgtgacgg cgtacctttt gtataatggg tcaacgactt 100020
aatttcagta gcaagcttaa ccatttaggg taggcgtagg gaaaccgagt cttaataggg 100080

cgctcagttg ctggaattag acccgaaacc gggatgatcta tccatggcca gtgtgaaggt 100140
cgagtaacat cgactggagg cgcgaaacca cttatgttga aaaatgaggg gatgagctgt 100200
ggataggagt gaaaggctaa tcaaaccggg agatagctgg ttctcttcga aaactattta 100260
ggtagtgcct cgtgtattac tgtagggggg agagcactgt ttcggctagg gggatcatccc 100320
gacttaccaa accgatgcaa actccgaata cctacaagta tgagcacggg agacagactg 100380
cgggtgctaa cgcccgtagt cgagagggaa acaaccaga ctgtcagcta aggtcccaaa 100440
ttatgattaa gtgggaaaca atgtgggaag gcacaaacag ctaggagggt ggcttagaag 100500
cagccatcct ttaaagaaag cgtaatagct cactagtcga gtcggcctgc gtggaagata 100560
taacggggct aaatcataaa ccgaagctac agatcttaaa tttatttaag atggtagaag 100620
agcgttctgt aagcggttga aggtaaagct agaggcgaac tggacgtatc agaagtgcga 100680
atgttgacat gagtaacgat caaagagggt aaaaacctct tcgccgaaaa accaagggtt 100740
cctgtccaac gctaatcgag gcagggtgag gcgccccccta aggcgagggc gaaagccgta 100800
gtcgatggga aacagggttaa tattcctgta ctttttataa ctgcgatggg gtgacggaga 100860
aggtagact agcacggcga cggttggtccg tgttcaaggt tgtaggctgg tgttctaggt 100920
aaatccggaa cgctaaggct gagaactgat aacgaccact ctacgagtgg gaagtagtcg 100980
ataccatgct tccaggaaaa acctctaagc ttcagggttat aagaaaccgt accctaaacc 101040
gacacagggt gttagggtcga gtagaccaag gtgtttgaga gaactatggg gaaggaaacta 101100
ggcaaaatag caccgtaact tcggggagaag gtgcgccgcg tttgggtgatg agacttgctc 101160
tctaagctga acgtgggtcga agataccagg tggctgcgac tgtttactaa aaacatagca 101220
ctctgcaaac tcgtaagagg aagtataggg tgtgacgcct gcccggtgcc ggaagggttaa 101280
ttgatggggg tagcttatgc gaagctcttg atcgaagccc cggtaaacgg cggccgtaac 101340
tataacgggt ctaaggtagc gaaattcctt gtcgggtaag ttccgacctg cacgaatggc 101400
gtaacgatgg ccacactgtc tccaccatag actcagtga attgaaatcg ctgttaagat 101460
gcagtgtacc cgcagctaga cggaaagacc ccgtgcacct ttactatagg ttcgcactgg 101520
actttgacct tacttggtga ggatagggtg gagactttga agcagagacg ccagtctttg 101580
tgtagtcac cttgaaatac cacccttgta agattgaagt tctaacctag gtccattatc 101640

tggatcaggg acagtgcgtg ctgggtagtt tgactggggc ggtctcctcc taaagagtaa 101700
cggaggagta cgaaggtatc cttatcacgg tcggacatcg tgaggtaagt ataaaggcag 101760
aaggatgctt gactgcgaga tcgacggatc gagcaggtag gaaactaggt cttagtgate 101820
cggtggttct gaatggaagg gccatcgctc aacggataaa aggtacgccg gggataacag 101880
gctgataccg cccaagagtt catatcgacg gcgggtgtttg gcacctcgat gtcggctcat 101940
cacatcctgg ggctggagca ggtcccaagg gtatggctgt tcgccattta aagtggtagc 102000
cgagctgggt ttagaacgtc gtgagacagt tcgggtcccta tctgctgtgg gcgtttggag 102060
at ttgagggga agctgattct agtacgagag gaccgaattg gacgaacctc tgggtgttccg 102120
gttgtcacgc cagtggcatt gccgggtagc tatgttcgga aaggataacc gctgaaagca 102180
tataagcggg aagcctctcc caagattaaa tctcccagag actttatgtc tcctaaagag 102240
tcgtcataga ctatgacgtt gataggcaag atgtgtaagc gctgcgaggc gttgagctaa 102300
cttgtactaa taactcgtga ggcttgatca tgtaacctta agcaagggtc ataatttgag 102360
taaaacattg tagtgagaat taaaaataa aaagttacat accagtttgc ctgatgacaa 102420
tagcaacttg gaaccacctg atcccatctc gaactcagaa gtgaaacgag ttaacgcaa 102480
tggtagtgcg ggtctccct gtgtgagagt aggaaatcgt caggcttttt tctttaaggc 102540
ttccagttta ctggaagcct ttttttttat ctcaagtata ataccagat gattattgga 102600
ttaacaggag gcattgggtc tggcaaatct gccgctgcag acttctttat tgatttaggt 102660
atatcagtct tagatgcaga tcaagttgct aaagaagctt tatctacaaa ttctcctgga 102720
tatactgatt ttatttctca atttggtgaa gtgtatttaa ataataatcg tgaggttgat 102780
aggctgaaat tgcgcgaact tattttttca aatccttcaa aaaaaagga tcttgagaat 102840
attattcatc ccatagttcg gtctgctatt agtaatttta ttattacatc aacatcacca 102900
tattctattg ttatgggtgcc actcattttt gaaacaaatt catataaaaa ttacgataag 102960
attattactg ttgactgtga tttagaactt cagatagtaa gggcctcaag tagagatgct 103020
caaaataaat cgcagattaa gaatattatt aataagcaag cctctagaga ggagaggcta 103080
agtatttctg atgacgtact tatcaataac agcaccctat ctgatctaaa aaaacaagtt 103140
aatgttttac atactaaata tatggagtta ttaaagagt agttgcccta gatgtgaaaa 103200

acctgtcaaa ctttctactg acaatattta tagacctttc tgctctgaga aatgcaaact 103260
tatagatttt ggtgattggg ctaatgaaga taataaaatt tcaagaccaa ttcaatctga 103320
agattttttac gaagattaaa ttttaagaaag tctccattca ccactttcaa tcatcggttt 103380
ggctttttta tacttcattt cttgcgtatc ttgaccgtta gtaattttta caagatcatt 103440
tctgcctaac ttaggttcgt ttcttgtcac agtacttggt ttactatag gagtggcttg 103500
tttttcattt tgaaagatat cagaattaat ttctctttt tccaatttta attcttgaga 103560
actattttgt ttgtttatgc tctctagctc agattctgtc gatatttgta atgagaaaag 103620
tattcttata gtttcaacat ctattttccga aagcatagat tcaaacattg aatatgcctc 103680
tcttttgaat tcafttttttg gatttttttg agcataagcc cttagaccaa cactatttct 103740
taaagatctt atttctgata aatgctcttt ccaatggaca tcaagtactt gcagcataac 103800
ctgcttctca agaagcaatc tattttcacc aagatcacta aacttttttg aatatttatt 103860
ttttgcttgc agaacaattt cttcggcaat tgtatttggt acaagttttt tattactatt 103920
aattttattg gctatatctg ttcttagtcc atagctctct tttaaataat catcgagctg 103980
cctacttttc cattgagact ctattgattc ttccaggtaca tataaattag atattccttt 104040
gaattgctgc tcgatgagtg actcaatggg actgctaata tcttctctct ctagtaattg 104100
attccttaaa gaatatatag cttgtctttg atcatttgat acatcgtcgt actctaaaag 104160
atttttcctt gcaccaaagt ttctgctctc aattcttttt tgcgcatttt ctattcctct 104220
tgaaagcatt ttagcttcaa tatgatcatc tcccatgccc agcctttcaa aagtagccct 104280
tcttccatcc gaaataaaaa gtctcaagag atcgtcttct aaagataaga agaattttga 104340
ataacctgga tcaccttgct tgccctgatct cccctttaac tgattatcta ttcttcttga 104400
ttcatgcctt tctgttccaa gtatatgaag cccaccagat tctataactt ttttattatt 104460
ttctttccac tctatatcgc ttgtgtcttc tttttttccg ccaagaacta tatccgtccc 104520
ccttctgccc atattcgtag caatagttac cattccaggt ttacctgcat tggcaattat 104580
ctctgcttct ttttcgtgat gcttggcatt taaaatttga tgaggatttt ttttgttatt 104640
taaagatgct gatacttctt ctgaagattc aaccgaaact gttccaacaa gaatgggaga 104700
agattttttt cttaattggt caatttcttc aattagagct ttatatttcg attctgttgt 104760

aagaaatact aagtcattaa gatcagctct agccatagga acatttggtg ggatgatgat 104820
 gacatttagg ccatagattt gactaaactc tactgcttca gtatctgctg ttctgtcat 104880
 cccagaaagt tttttaaata atctaaaaaa gttttggaat gtggtggatg ctagtgtttg 104940
 agactctctt tggatagcaa cattttcttt gcattccagt gcctggtgaa caccttcact 105000
 cattcttctt cggggcattg ttctacctgt atgctcatca atcaaaagaa cctcaccgtt 105060
 cctaaccaa taatccacat tctttttaaa taagaagctt gctctaagtg ttgcttgaac 105120
 aaatttcata atttttaaat tagaaacaga gtaagcccat ctgaggctcc aagccgattc 105180
 cagc 105184

<210> 2

<211> 29

<212> DNA

<213> pcr primer

<220>

<221> oligo-nucleotide-primer

<222> (1)..(29)

<300>

<301> Beja, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A.,
 Nguyen, L.P., Jovanovich, S.B., Gates, C.M., Feldman, R.A., Spudich, J.L.,
 Spudich, E.N. and DeLong, E.F.

<302> Bacterial rhodopsin: evidence for a new type of phototrophy in the sea

<303> Science

<304> 289

<305> 5486

<306> 1902-1906

<307> 2000-09-15

<309> ____-__-__

<400> 2

accatgggta aattattact gatattagg

29

<210> 3

<211> 24

<212> DNA

<213> pcr primer

<220>

<221> oligo-nucleotide-primer

<222> (1)..(24)

<300>

<301> Beja,O., Aravind,L., Koonin,E.V., Suzuki,M.T., Hadd,A.,
 Nguyen,L.P., Jovanovich,S.B., Gates,C.M., Feldman,R.A., Spudich,J.L.,
 Spudich,E.N. and DeLong,E.F.

<302> Bacterial rhodopsin: evidence for a new type of phototrophy in the
 sea

<303> Science

<304> 289

<305> 5486

<306> 1902-1906

<307> 2000-09-15

<309> _____

<400> 3

agcattagaa gattctttaa cagc

24

<210> 4

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> light-driven proton pump; has the properties of a light-driven pr
 oton pump when expressed with retinal in Escherichia coli. Note
 that additional three nucleotide residues incorporated by pcr pri
 ming with reference to the original 31A08 DNA sequence (DNA resid
 ues 4-6, ggt), adding a new restriction site for cloning

<300>

<301> Beja,O., Aravind,L., Koonin,E.V., Suzuki,M.T.,
 Hadd,A.,Nguyen,L.P., Jovanovich,S.B., Gates,C.M.,
 Feldman,R.A.,Spudich,J.L., Spudich,E.N. and DeLong,E.F

<302> Bacterial rhodopsin: evidence for a new type of phototrophy in the
 sea

<303> Science

<304> 289

<305> 5486

<306> 1902-1906

<307> 2000-09-15

<308> AAG10475

<309> 2000-06-15

<313> (1)..(750)

<400> 4

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

48

ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val 20 25 30	96
tct ttt tgg tta gtt act gct gct tta tta gca tct act gta ttt ttc Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe 35 40 45	144
ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr 50 55 60	192
gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met 65 70 75 80	240
aga ggg gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr 85 90 95	288
att gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu 100 105 110	336
att ctt gct gct gca act aat gtt gct gga tca tta ttt aag aaa tta Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu 115 120 125	384
cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala 130 135 140	432
gga atc atg gct gca tgg cct gca ttc att att ggg tgt tta gct tgg Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp 145 150 155 160	480
gta tac atg att tat gaa tta tgg gct gga gaa gga aaa tct gca tgt Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys 165 170 175	528
aat act gca agt cct gct gtg caa tca gct tac aac aca atg atg tat Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr 180 185 190	576
att atc atc ttt ggt tgg gcg att tat cct gta ggt tat ttc aca ggt Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly 195 200 205	624
tac ctg atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr 210 215 220	672

aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720
 Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240

aat gtt gct gtt aaa gaa tct tct aat gct 750
 Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 5
 <211> 250
 <212> PRT
 <213> Naturally occurring gamma proteobacterium

<400> 5

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
 100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
 115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
 130 135 140

Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
 145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys
 165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
 180 185 190

Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
 210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 6
 <211> 747
 <212> DNA
 <213> Naturally occurring gamma proteobacterium

<220>
 <221> CDS
 <222> (1)..(747)
 <223> Native proteorhodopsin DNA sequence from BAC clone 31A08

<300>
 <301> Beja, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A.,
 Nguyen, L.P., Jovanovich, S.B., Gates, C.M., Feldman, R.A., Spudich, J.L.,
 Spudich, E.N. and DeLong, E.F.
 <302> Bacterial rhodopsin: evidence for a new type of phototrophy in the
 sea
 <303> Science
 <304> 289
 <305> 5486
 <306> 1902-1906
 <307> 2000-09-15
 <309> _____

<313> (1) .. (747)

<400> 6

atg	aaa	tta	tta	ctg	ata	tta	ggt	agt	ggt	att	gca	ctt	cct	aca	ttt	48
Met	Lys	Leu	Leu	Leu	Ile	Leu	Gly	Ser	Val	Ile	Ala	Leu	Pro	Thr	Phe	
1				5					10					15		
gct	gca	ggt	ggt	ggt	gac	ctt	gat	gct	agt	gat	tac	act	ggt	ggt	tct	96
Ala	Ala	Gly	Gly	Gly	Asp	Leu	Asp	Ala	Ser	Asp	Tyr	Thr	Gly	Val	Ser	
		20						25					30			
ttt	tgg	tta	ggt	act	gct	gct	tta	tta	gca	tct	act	gta	ttt	ttc	ttt	144
Phe	Trp	Leu	Val	Thr	Ala	Ala	Leu	Leu	Ala	Ser	Thr	Val	Phe	Phe	Phe	
		35					40					45				
ggt	gaa	aga	gat	aga	ggt	tct	gca	aaa	tgg	aaa	aca	tca	tta	act	gta	192
Val	Glu	Arg	Asp	Arg	Val	Ser	Ala	Lys	Trp	Lys	Thr	Ser	Leu	Thr	Val	
	50					55					60					
tct	ggt	ctt	ggt	act	ggt	att	gct	ttc	tgg	cat	tac	atg	tac	atg	aga	240
Ser	Gly	Leu	Val	Thr	Gly	Ile	Ala	Phe	Trp	His	Tyr	Met	Tyr	Met	Arg	
65					70				75					80		
ggg	gta	tgg	att	gaa	act	ggt	gat	tcg	cca	act	gta	ttt	aga	tac	att	288
Gly	Val	Trp	Ile	Glu	Thr	Gly	Asp	Ser	Pro	Thr	Val	Phe	Arg	Tyr	Ile	
				85					90				95			
gat	tgg	tta	cta	aca	ggt	cct	cta	tta	ata	tgt	gaa	ttc	tac	tta	att	336
Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Leu	Ile	Cys	Glu	Phe	Tyr	Leu	Ile	
		100					105						110			
ctt	gct	gct	gca	act	aat	ggt	gct	gga	tca	tta	ttt	aag	aaa	tta	cta	384
Leu	Ala	Ala	Ala	Thr	Asn	Val	Ala	Gly	Ser	Leu	Phe	Lys	Lys	Leu	Leu	
	115						120					125				
ggt	ggt	tct	ctt	ggt	atg	ctt	gtg	ttt	ggt	tac	atg	ggt	gaa	gca	gga	432
Val	Gly	Ser	Leu	Val	Met	Leu	Val	Phe	Gly	Tyr	Met	Gly	Glu	Ala	Gly	
	130					135					140					
atc	atg	gct	gca	tgg	cct	gca	ttc	att	att	ggg	tgt	tta	gct	tgg	gta	480
Ile	Met	Ala	Ala	Trp	Pro	Ala	Phe	Ile	Ile	Gly	Cys	Leu	Ala	Trp	Val	
145					150					155					160	
tac	atg	att	tat	gaa	tta	tgg	gct	gga	gaa	gga	aaa	tct	gca	tgt	aat	528
Tyr	Met	Ile	Tyr	Glu	Leu	Trp	Ala	Gly	Glu	Gly	Lys	Ser	Ala	Cys	Asn	
				165				170					175			
act	gca	agt	cct	gct	gtg	caa	tca	gct	tac	aac	aca	atg	atg	tat	att	576
Thr	Ala	Ser	Pro	Ala	Val	Gln	Ser	Ala	Tyr	Asn	Thr	Met	Met	Tyr	Ile	
			180					185					190			
atc	atc	ttt	ggt	tgg	gcg	att	tat	cct	gta	ggt	tat	ttc	aca	ggt	tac	624

Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly Tyr
 195 200 205

ctg atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat aac 672
 Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr Asn
 210 215 220

ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg aat 720
 Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp Asn
 225 230 235 240

gtt gct gtt aaa gaa tct tct aat gct 747
 Val Ala Val Lys Glu Ser Ser Asn Ala
 245

<210> 7
 <211> 249
 <212> PRT
 <213> Naturally occurring gamma proteobacterium
 <400> 7

Met Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr Phe
 1 5 10 15

Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val Ser
 20 25 30

Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe Phe
 35 40 45

Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr Val
 50 55 60

Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met Arg
 65 70 75 80

Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr Ile
 85 90 95

Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu Ile
 100 105 110

Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu Leu

115 120 125
 Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala Gly
 130 135 140
 Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp Val
 145 150 155 160
 Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys Asn
 165 170 175
 Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr Ile
 180 185 190
 Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly Tyr
 195 200 205
 Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr Asn
 210 215 220
 Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp Asn
 225 230 235 240
 Val Ala Val Lys Glu Ser Ser Asn Ala
 245

<210> 8
 <211> 750
 <212> DNA
 <213> Naturally occurring gamma proteobacterium

<220>
 <221> CDS
 <222> (1)..(750)
 <223> proteorhodopsin variant from clone EBAC40

<400> 8
 atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct acā 48
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15
 ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt 96

Phe	Ala	Ala	Gly	Gly	Gly	Asp	Leu	Asp	Ala	Ser	Asp	Tyr	Thr	Gly	Val		
			20					25					30				
tct	ttt	tgg	tta	gtt	act	gct	gct	cta	tta	gca	tct	act	gta	ttt	ttc		144
Ser	Phe	Trp	Leu	Val	Thr	Ala	Ala	Leu	Leu	Ala	Ser	Thr	Val	Phe	Phe		
		35					40					45					
ttt	gtt	gaa	aga	gat	aga	gtt	tct	gca	aaa	tgg	aaa	aca	tca	tta	act		192
Phe	Val	Glu	Arg	Asp	Arg	Val	Ser	Ala	Lys	Trp	Lys	Thr	Ser	Leu	Thr		
	50					55					60						
gta	tcg	ggg	ctt	gtt	act	ggg	att	gct	ttc	tgg	cat	tac	atg	tac	atg		240
Val	Ser	Gly	Leu	Val	Thr	Gly	Ile	Ala	Phe	Trp	His	Tyr	Met	Tyr	Met		
65					70					75					80		
aga	ggg	gta	tgg	att	gag	act	ggg	gat	tcg	cca	act	gta	ttt	aga	tac		288
Arg	Gly	Val	Trp	Ile	Glu	Thr	Gly	Asp	Ser	Pro	Thr	Val	Phe	Arg	Tyr		
				85					90					95			
att	gat	tgg	tta	cta	aca	gtt	cct	cta	ttg	ata	tgt	gaa	ttc	tac	tta		336
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Leu	Ile	Cys	Glu	Phe	Tyr	Leu		
			100					105					110				
att	ctt	gct	gct	gca	aca	aat	gtt	gct	gct	ggc	ctg	ttt	aag	aaa	tta		384
Ile	Leu	Ala	Ala	Ala	Thr	Asn	Val	Ala	Ala	Gly	Leu	Phe	Lys	Lys	Leu		
		115					120					125					
ttg	gtt	ggg	tct	ctt	gtt	atg	ctt	gtg	ttt	ggg	tac	atg	ggg	gag	gca		432
Leu	Val	Gly	Ser	Leu	Val	Met	Leu	Val	Phe	Gly	Tyr	Met	Gly	Glu	Ala		
	130					135					140						
gga	att	atg	aac	gct	tgg	ggg	gca	ttc	gtt	att	ggg	tgt	tta	gct	tgg		480
Gly	Ile	Met	Asn	Ala	Trp	Gly	Ala	Phe	Val	Ile	Gly	Cys	Leu	Ala	Trp		
145					150					155					160		
gta	tac	atg	att	tat	gaa	cta	tgg	gct	gga	gaa	ggc	aag	gct	gca	tgt		528
Val	Tyr	Met	Ile	Tyr	Glu	Leu	Trp	Ala	Gly	Glu	Gly	Lys	Ala	Ala	Cys		
				165					170					175			
aat	act	gca	agt	cct	gct	gtg	caa	tca	gct	tac	aac	aca	atg	atg	tat		576
Asn	Thr	Ala	Ser	Pro	Ala	Val	Gln	Ser	Ala	Tyr	Asn	Thr	Met	Met	Tyr		
			180					185					190				
ata	atc	atc	ttt	ggg	tgg	gca	att	tat	cct	gta	ggg	tat	ttc	aca	ggg		624
Ile	Ile	Ile	Phe	Gly	Trp	Ala	Ile	Tyr	Pro	Val	Gly	Tyr	Phe	Thr	Gly		
		195					200					205					
tac	cta	atg	ggg	gac	ggg	gga	tca	gct	ctt	aac	tta	aac	ctt	atc	tat		672
Tyr	Leu	Met	Gly	Asp	Gly	Gly	Ser	Ala	Leu	Asn	Leu	Asn	Leu	Ile	Tyr		
	210					215					220						
gac	ctt	gct	gac	ttt	gtt	aac	aag	att	cta	ttt	ggg	tta	att	ata	tgg		720

Asp Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240

aat gtt gct gtt aaa gaa tct tct aat gct
 Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

750

<210> 9
 <211> 250
 <212> PRT
 <213> Naturally occurring gamma proteobacterium

<400> 9

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
 100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu
 115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
 130 135 140

Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp

145 150 155 160
 Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ala Ala Cys
 165 170 175
 Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
 180 185 190
 Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
 210 215 220
 Asp Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240
 Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 10
 <211> 750
 <212> DNA
 <213> Naturally occurring gamma proteobacterium

<220>
 <221> CDS
 <222> (1)..(750)
 <223> proteorhodopsin variant from clone EBAC41

<400> 10
 atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca 48
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15
 ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt 96
 Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30
 tct ttt tgg tta gct act gct gct tta tta gca tct act gta ttt ttc 144
 Ser Phe Trp Leu Ala Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45
 ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192

Phe	Val	Glu	Arg	Asp	Arg	Val	Ser	Ala	Lys	Trp	Lys	Thr	Ser	Leu	Thr		
50						55					60						
gta	tct	ggt	ctt	gtt	act	ggt	att	gct	ttc	tgg	cat	tac	atg	tac	atg	240	
Val	Ser	Gly	Leu	Val	Thr	Gly	Ile	Ala	Phe	Trp	His	Tyr	Met	Tyr	Met		
65					70					75					80		
aga	ggg	gta	tgg	att	gaa	act	ggt	gat	tcg	cca	act	gta	ttt	aga	tac	288	
Arg	Gly	Val	Trp	Ile	Glu	Thr	Gly	Asp	Ser	Pro	Thr	Val	Phe	Arg	Tyr		
				85					90					95			
att	gat	tgg	tta	cta	aca	gtt	cct	cta	tta	ata	tgt	gaa	ttc	tac	tta	336	
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Leu	Ile	Cys	Glu	Phe	Tyr	Leu		
			100					105					110				
att	ctt	gct	gct	gct	act	aat	gtt	gct	gga	tca	tta	ttt	aag	aaa	tta	384	
Ile	Leu	Ala	Ala	Ala	Thr	Asn	Val	Ala	Gly	Ser	Leu	Phe	Lys	Lys	Leu		
		115					120					125					
cta	gtt	ggt	tct	ctt	gtt	atg	ctt	gtg	ttt	ggt	tac	atg	ggt	gaa	gca	432	
Leu	Val	Gly	Ser	Leu	Val	Met	Leu	Val	Phe	Gly	Tyr	Met	Gly	Glu	Ala		
	130					135					140						
gga	atc	atg	gct	gca	tgg	cct	gca	ttc	att	att	ggg	tgt	tta	gct	tgg	480	
Gly	Ile	Met	Ala	Ala	Trp	Pro	Ala	Phe	Ile	Ile	Gly	Cys	Leu	Ala	Trp		
145					150					155					160		
gta	tac	atg	att	tat	gaa	cta	tgg	gct	gga	gaa	gga	aaa	tct	gca	tgt	528	
Val	Tyr	Met	Ile	Tyr	Glu	Leu	Trp	Ala	Gly	Glu	Gly	Lys	Ser	Ala	Cys		
				165					170					175			
aat	act	gca	agt	cct	gct	gtg	caa	tca	gct	tac	aac	aca	atg	atg	tat	576	
Asn	Thr	Ala	Ser	Pro	Ala	Val	Gln	Ser	Ala	Tyr	Asn	Thr	Met	Met	Tyr		
			180					185					190				
att	atc	atc	ttt	ggt	tgg	gcg	att	tat	cct	gta	ggt	tat	ttc	aca	ggt	624	
Ile	Ile	Ile	Phe	Gly	Trp	Ala	Ile	Tyr	Pro	Val	Gly	Tyr	Phe	Thr	Gly		
		195					200					205					
tac	ctg	atg	ggt	gac	ggt	gga	tca	gct	ctt	aac	tta	aac	ctt	atc	tat	672	
Tyr	Leu	Met	Gly	Asp	Gly	Gly	Ser	Ala	Leu	Asn	Leu	Asn	Leu	Ile	Tyr		
	210					215					220						
aac	ctt	gct	gat	ttt	gtt	aac	aag	att	cta	ttt	ggt	tta	att	ata	tgg	720	
Asn	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile	Trp		
225					230					235					240		
aat	gtt	gct	gtt	aaa	gaa	tct	tct	aat	gct							750	
Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala								
				245					250								

<210> 11
 <211> 250
 <212> PRT
 <213> Naturally occurring gamma proteobacterium

<400> 11

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30

Ser Phe Trp Leu Ala Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
 100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
 115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
 130 135 140

Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
 145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys
 165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr

180

185

190

Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
 210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 12
 <211> 750
 <212> DNA
 <213> Naturally occurring gamma proteobacterium

<220>
 <221> CDS
 <222> (1)..(750)
 <223> Proteorhodopsin variant from clone EBAC64

<400> 12
 atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca 48
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15
 ttt gct gca ggt ggc ggt gac ctt gat gct agt gat tac act ggt gtt 96
 Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30
 tct ttt tgg tta gtt aca gct gct cta tta gca tct act gta ttt ttc 144
 Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45
 ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192
 Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60
 gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg 240
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80
 aga gga gta tgg att gaa act ggt gat tcg cct act gta ttt aga tac 288

Arg	Gly	Val	Trp	Ile	Glu	Thr	Gly	Asp	Ser	Pro	Thr	Val	Phe	Arg	Tyr		
				85					90					95			
att	gat	tgg	tta	cta	aca	gtt	cct	tta	tta	ata	tgt	gaa	ttc	tac	tta		336
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Leu	Ile	Cys	Glu	Phe	Tyr	Leu		
			100					105					110				
att	ctt	gct	gct	gca	act	aat	gtt	gcc	ggc	tca	tta	ttt	aag	aaa	ctt		384
Ile	Leu	Ala	Ala	Ala	Thr	Asn	Val	Ala	Gly	Ser	Leu	Phe	Lys	Lys	Leu		
		115					120					125					
cta	gtt	ggg	tct	ctt	gtt	atg	ctt	gtg	ttt	ggg	tac	atg	ggg	gaa	gca		432
Leu	Val	Gly	Ser	Leu	Val	Met	Leu	Val	Phe	Gly	Tyr	Met	Gly	Glu	Ala		
		130				135					140						
gga	att	atg	gca	gct	tgg	cct	gca	ttc	att	att	ggg	tgt	tta	gct	tgg		480
Gly	Ile	Met	Ala	Ala	Trp	Pro	Ala	Phe	Ile	Ile	Gly	Cys	Leu	Ala	Trp		
145					150				155					160			
gta	tac	atg	att	tat	gaa	cta	tat	gct	gga	gaa	gga	aaa	tct	gca	tgt		528
Val	Tyr	Met	Ile	Tyr	Glu	Leu	Tyr	Ala	Gly	Glu	Gly	Lys	Ser	Ala	Cys		
				165					170					175			
aat	act	gca	agt	cct	tcg	gtt	caa	tca	gct	tac	aac	aca	atg	atg	gct		576
Asn	Thr	Ala	Ser	Pro	Ser	Val	Gln	Ser	Ala	Tyr	Asn	Thr	Met	Met	Ala		
			180					185					190				
atc	ata	gtc	ttc	ggg	tgg	gca	att	tat	cct	ata	ggg	tat	ttc	aca	ggg		624
Ile	Ile	Val	Phe	Gly	Trp	Ala	Ile	Tyr	Pro	Ile	Gly	Tyr	Phe	Thr	Gly		
		195				200						205					
tac	cta	atg	ggg	gac	ggg	gga	tca	gct	ctt	aac	tta	aac	ctt	att	tat		672
Tyr	Leu	Met	Gly	Asp	Gly	Gly	Ser	Ala	Leu	Asn	Leu	Asn	Leu	Ile	Tyr		
		210				215					220						
aac	ctt	gct	gac	ttt	gtt	aac	aag	att	cta	ttt	ggg	tta	att	ata	tgg		720
Asn	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile	Trp		
225					230					235				240			
aat	gtt	gct	gtt	aaa	gaa	tct	tct	aat	gct								750
Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala								
				245					250								

<210> 13

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 13

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr

1	5	10	15
Phe	Ala	Ala	Gly
	20		
Gly	Gly	Gly	Asp
			25
Leu	Asp	Ala	Ser
			30
Thr	Gly	Val	
Ser	Phe	Trp	Leu
	35		
Val	Thr	Ala	Ala
			40
Leu	Leu	Ala	Ser
			45
Thr	Val	Phe	Phe
Phe	Val	Glu	Arg
	50		
Asp	Arg	Val	Ser
			55
Ala	Lys	Trp	Lys
			60
Thr	Ser	Leu	Thr
Val	Ser	Gly	Leu
	65		
Val	Thr	Gly	Ile
			70
Ala	Phe	Trp	His
			75
Tyr	Met	Tyr	Met
			80
Arg	Gly	Val	Trp
			85
Ile	Glu	Thr	Gly
			90
Asp	Ser	Pro	Thr
			95
Val	Phe	Arg	Tyr
Ile	Asp	Trp	Leu
			100
Leu	Thr	Val	Pro
			105
Leu	Ile	Cys	Glu
			110
Phe	Tyr	Leu	
Ile	Leu	Ala	Ala
	115		
Ala	Thr	Asn	Val
			120
Ala	Gly	Ser	Leu
			125
Phe	Lys	Lys	Leu
Leu	Val	Gly	Ser
	130		
Leu	Val	Met	Leu
			135
Val	Phe	Gly	Tyr
			140
Met	Gly	Glu	Ala
Gly	Ile	Met	Ala
	145		
Ala	Trp	Pro	Ala
			150
Phe	Ile	Ile	Gly
			155
Cys	Leu	Ala	Trp
			160
Val	Tyr	Met	Ile
			165
Tyr	Glu	Leu	Tyr
			170
Ala	Gly	Glu	Gly
			175
Lys	Ser	Ala	Cys
Asn	Thr	Ala	Ser
			180
Pro	Ser	Val	Gln
			185
Ser	Ala	Tyr	Asn
			190
Thr	Met	Met	Ala
Ile	Ile	Val	Phe
			195
Gly	Trp	Ala	Ile
			200
Tyr	Pro	Ile	Gly
			205
Tyr	Phe	Thr	Gly
Tyr	Leu	Met	Gly
Asp	Gly	Gly	Ser
Ala	Leu	Asn	Leu
Asn	Leu	Ile	Tyr

210

215

220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 14

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> Proteorhodopsin variant from pcr clone HOT01m: GenBank# AF349978

<400> 14

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca 48
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt 96
 Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30

tct ttt tgg tta gtt act gct gct cta tta gca tct act gta ttt ttc 144
 Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45

ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192
 Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

gta tcg ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg 240
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80

aga ggg gta tgg att gag acc ggt gat tcg cca act gta ttt aga tac 288
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95

att gat tgg tta cta aca gtt cct cta ttg ata tgt gaa ttc tac tta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
 100 105 110

att ctt gct gct gca aca aat gtt gct gct ggc ctg ttt aag aaa tta 384
 Ile Leu Ala Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu
 115 120 125

ttg gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gag gca 432
 Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
 130 135 140

gga att atg aac gct tgg ggt gca ttc gtt att ggg tgt tta gct tgg 480
 Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp
 145 150 155 160

gta tac atg att tat gaa cta tgg gct gga gaa ggc aag gct gca tgt 528
 Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ala Ala Cys
 165 170 175

aat act gca agt cct gct gtg caa tca gct tac aac aca atg atg tat 576
 Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
 180 185 190

ata atc atc ttt ggt tgg gca att tat cct gta ggt tat ttc aca ggt 624
 Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205

tac cta atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat 672
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
 210 215 220

aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720
 Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240

aat gtt gct gtt aaa gaa tct tct aat gct 750
 Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 15

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 15

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu
115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp
145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ala Ala Cys
165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
180 185 190

Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 16
 <211> 753
 <212> DNA
 <213> Naturally occurring gamma prtoeobacterium
 <220>
 <221> CDS
 <222> (1)..(753)
 <223> Proteorhodopsin variant from pcr clone HOT75m1: GenBank#AF349979

<400> 16
 atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15
 ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
 Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30
 tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gtg ttc ttt 144
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
 35 40 45
 ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt gct 192
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Ala
 50 55 60
 gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg 240
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80
 aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat 288
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95
 att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
 100 105 110
 att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt 384
 Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
 115 120 125
 cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct 432
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
 130 135 140

gga tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg 480
 Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
 145 150 155 160

 tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta 528
 Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
 165 170 175

 agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg atg 576
 Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met
 180 185 190

 att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt 624
 Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
 195 200 205

 tac cta atg ggt ggc gaa ggt gta tac gct tca aac tta aac ctt ata 672
 Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
 210 215 220

 tat aac ctt gcc gac ctt gtt aac aag att cta ttt ggt ttg atc att 720
 Tyr Asn Leu Ala Asp Leu Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
 225 230 235 240

 tgg aat gtt gct gtt aaa gaa tct tct aat gct 753
 Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 17
 <211> 251
 <212> PRT
 <213> Naturally occurring gamma prtoeobacterium

 <400> 17

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Ala
 50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
115 120 125

Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
130 135 140

Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
145 150 155 160

Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
165 170 175

Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met
180 185 190

Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
195 200 205

Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
210 215 220

Tyr Asn Leu Ala Asp Leu Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 18
<211> 753
<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone HOT75m3; GenBank#AF349980

<400> 18

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca	48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser	
1 5 10 15	
ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt	96
Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val	
20 25 30	
tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gta ttc ttt	144
Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe	
35 40 45	
ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act	192
Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	
gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tac atg	240
Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met	
65 70 75 80	
aga ggt gtt tgg ata gat act ggt gat aca cca aca gta ttt aga tat	288
Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr	
85 90 95	
att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta	336
Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu	
100 105 110	
att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt	384
Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu	
115 120 125	
cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct	432
Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala	
130 135 140	
ggt tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg	480
Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp	
145 150 155 160	
tta tac atg att tat gag cta cat atg ggt gaa ggt aag gct gct gta	528
Leu Tyr Met Ile Tyr Glu Leu His Met Gly Glu Gly Lys Ala Ala Val	

165										170					175					
agt	act	gca	agt	cct	gct	gtt	aac	tct	gca	tac	aac	gca	atg	atg	aag	576				
Ser	Thr	Ala	Ser	Pro	Ala	Val	Asn	Ser	Ala	Tyr	Asn	Ala	Met	Met	Lys					
			180				185				190									
att	att	gtt	att	gga	tgg	gca	att	tat	cct	gct	gga	tat	gct	gct	ggt	624				
Ile	Ile	Val	Ile	Gly	Trp	Ala	Ile	Tyr	Pro	Ala	Gly	Tyr	Ala	Ala	Gly					
			195				200				205									
tac	cta	atg	agt	ggt	gac	ggt	gta	tac	gct	tca	aac	tta	aac	ctt	ata	672				
Tyr	Leu	Met	Ser	Gly	Asp	Gly	Val	Tyr	Ala	Ser	Asn	Leu	Asn	Leu	Ile					
			210				215				220									
tat	aac	ctt	gct	gac	ttt	gtt	aac	aag	att	cta	ttt	ggt	ttg	atc	att	720				
Tyr	Asn	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile					
			225				230				235				240					
tgg	aat	gtt	gct	gtt	aaa	gaa	tct	tct	aat	gct						753				
Trp	Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala										
			245				250													

<210> 19

<211> 251

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 19

Met	Gly	Lys	Leu	Leu	Leu	Ile	Leu	Gly	Ser	Ala	Ile	Ala	Leu	Pro	Ser
1				5					10					15	

Phe	Ala	Ala	Ala	Gly	Gly	Asp	Leu	Asp	Ile	Ser	Asp	Thr	Val	Gly	Val
			20					25					30		

Ser	Phe	Trp	Leu	Val	Thr	Ala	Gly	Met	Leu	Ala	Ala	Thr	Val	Phe	Phe
		35					40					45			

Phe	Val	Glu	Arg	Asp	Gln	Val	Ser	Ala	Lys	Trp	Lys	Thr	Ser	Leu	Thr
	50					55					60				

Val	Ser	Gly	Leu	Ile	Thr	Gly	Ile	Ala	Phe	Trp	His	Tyr	Leu	Tyr	Met
65					70					75				80	

Arg	Gly	Val	Trp	Ile	Asp	Thr	Gly	Asp	Thr	Pro	Thr	Val	Phe	Arg	Tyr
				85					90					95	

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
115 120 125

Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
130 135 140

Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
145 150 155 160

Leu Tyr Met Ile Tyr Glu Leu His Met Gly Glu Gly Lys Ala Ala Val
165 170 175

Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Lys
180 185 190

Ile Ile Val Ile Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
195 200 205

Tyr Leu Met Ser Gly Asp Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
210 215 220

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 20

<211> 753

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone HOT75m4; GenBank #AF349981

<400> 20
 atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15

ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
 Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30

tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gtg ttc ttt 144
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
 35 40 45

ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act 192
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg 240
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80

aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat 288
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95

att gat tgg tta tta act gtt cca tta caa gtg gtt gag ttc tat cta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Val Val Glu Phe Tyr Leu
 100 105 110

att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt 384
 Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
 115 120 125

cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct 432
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
 130 135 140

gga tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg 480
 Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
 145 150 155 160

tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta 528
 Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
 165 170 175

agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg atg 576
 Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met
 180 185 190

att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt 624
 Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly

195	200	205	
tac cta atg ggt ggc gaa ggt gta tac gct tca aac tta aac ctt ata			672
Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile			
210	215	220	
tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att			720
Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile			
225	230	235	240
tgg aat gtt gct gtt aaa gaa tct tct aat gct			753
Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala			
245	250		

<210> 21
 <211> 251
 <212> PRT
 <213> Naturally occurring gamma proteobacterium
 <400> 21

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Val Val Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
115 120 125

Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
130 135 140

Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
145 150 155 160

Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
165 170 175

Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met
180 185 190

Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
195 200 205

Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
210 215 220

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 22

<211> 753

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone HOT75m8: GenBank#AF349982

<400> 22

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
1 5 10 15

ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96

Phe	Ala	Ala	Ala	Gly	Gly	Asp	Leu	Asp	Ile	Ser	Asp	Thr	Val	Gly	Val		
			20					25					30				
tca	ttc	tgg	ctg	gtt	aca	gct	ggg	atg	tta	gcg	gca	act	gtg	ttc	ttt		144
Ser	Phe	Trp	Leu	Val	Thr	Ala	Gly	Met	Leu	Ala	Ala	Thr	Val	Phe	Phe		
		35					40					45					
ttt	gta	gaa	aga	gac	caa	gtc	agc	gct	aag	tgg	aaa	act	tca	ctt	act		192
Phe	Val	Glu	Arg	Asp	Gln	Val	Ser	Ala	Lys	Trp	Lys	Thr	Ser	Leu	Thr		
	50					55					60						
gta	tct	ggg	tta	att	act	ggg	ata	gct	ttt	tgg	cat	tat	ctc	tat	atg		240
Val	Ser	Gly	Leu	Ile	Thr	Gly	Ile	Ala	Phe	Trp	His	Tyr	Leu	Tyr	Met		
65					70					75					80		
aga	ggg	gtt	tgg	ata	gac	act	ggg	gat	acc	cca	aca	gta	ttc	aga	tat		288
Arg	Gly	Val	Trp	Ile	Asp	Thr	Gly	Asp	Thr	Pro	Thr	Val	Phe	Arg	Tyr		
				85					90					95			
att	gat	tgg	tta	tta	act	gtt	cca	tta	caa	atg	gtt	gag	ttc	tat	cta		336
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Gln	Met	Val	Glu	Phe	Tyr	Leu		
			100					105					110				
att	ctt	gct	gct	tgt	aca	aat	gtt	gct	gct	tca	tta	ttt	aag	aag	ctt		384
Ile	Leu	Ala	Ala	Cys	Thr	Asn	Val	Ala	Ala	Ser	Leu	Phe	Lys	Lys	Leu		
		115					120					125					
cta	gct	ggg	tca	tta	gta	atg	tta	ggg	gct	gga	ttt	gca	ggc	gaa	gct		432
Leu	Ala	Gly	Ser	Leu	Val	Met	Leu	Gly	Ala	Gly	Phe	Ala	Gly	Glu	Ala		
	130					135					140						
gga	ttg	gct	cct	gta	tgg	cct	gct	ttc	att	att	ggg	atg	gct	gga	tgg		480
Gly	Leu	Ala	Pro	Val	Trp	Pro	Ala	Phe	Ile	Ile	Gly	Met	Ala	Gly	Trp		
145					150					155					160		
tta	tac	atg	att	tat	gag	cta	tat	atg	ggg	gaa	ggg	aag	gct	gct	gta		528
Leu	Tyr	Met	Ile	Tyr	Glu	Leu	Tyr	Met	Gly	Glu	Gly	Lys	Ala	Ala	Val		
				165					170					175			
agt	act	gca	agt	cct	gct	gtt	aac	tct	gca	tac	aac	gca	atg	atg	gtg		576
Ser	Thr	Ala	Ser	Pro	Ala	Val	Asn	Ser	Ala	Tyr	Asn	Ala	Met	Met	Val		
			180					185					190				
att	att	gtt	gtt	gga	tgg	gca	att	tat	cct	gct	gga	tat	gct	gct	ggg		624
Ile	Ile	Val	Val	Gly	Trp	Ala	Ile	Tyr	Pro	Ala	Gly	Tyr	Ala	Ala	Gly		
		195				200						205					
tac	cta	atg	ggg	ggc	gaa	ggg	gta	tac	gct	tca	aac	tta	aac	ctt	ata		672
Tyr	Leu	Met	Gly	Gly	Glu	Gly	Val	Tyr	Ala	Ser	Asn	Leu	Asn	Leu	Ile		
	210					215					220						
tat	aac	ctt	gcc	gac	ctt	gtt	aac	aag	att	cta	ttt	ggg	ttg	atc	att		720

Tyr Asn Leu Ala Asp Leu Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
 225 230 235 240

tgg aat gtt gct gtt aaa gaa tct tct aat gct
 Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

753

<210> 23
 <211> 251
 <212> PRT
 <213> Naturally occurring gamma proteobacterium

<400> 23

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
 100 105 110

Ile Leu Ala Ala Cys Thr Asn Val Ala Ala Ser Leu Phe Lys Lys Leu
 115 120 125

Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
 130 135 140

Gly Leu Ala Pro Val Trp Pro Ala Phe Ile Ile Gly Met Ala Gly Trp

145 150 155 160

Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
165 170 175

Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Val
180 185 190

Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
195 200 205

Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
210 215 220

Tyr Asn Leu Ala Asp Leu Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 24

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> Proteorhodopsin variant from pcr clone MB0m1: GenBank#AF349983

<400> 24

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca 48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1 5 10 15

ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt 96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20 25 30

tct ttt tgg tta gtt act gct gct cta tta gca tct act gta ttt ttc 144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45

ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192

Phe	Val	Glu	Arg	Asp	Arg	Val	Ser	Ala	Lys	Trp	Lys	Thr	Ser	Leu	Thr		
50						55					60						
gta	tct	ggt	ctt	gtt	act	ggt	att	gct	ttc	tgg	cat	tac	atg	tac	atg	240	
Val	Ser	Gly	Leu	Val	Thr	Gly	Ile	Ala	Phe	Trp	His	Tyr	Met	Tyr	Met		
65					70					75					80		
aga	ggg	gta	tgg	att	gag	act	ggt	gat	tcg	cca	act	gta	ttt	aga	tac	288	
Arg	Gly	Val	Trp	Ile	Glu	Thr	Gly	Asp	Ser	Pro	Thr	Val	Phe	Arg	Tyr		
				85					90					95			
att	gat	tgg	tta	cta	aca	gtt	cct	cta	ttg	ata	tgt	gaa	ttc	tac	tta	336	
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Leu	Ile	Cys	Glu	Phe	Tyr	Leu		
			100					105					110				
att	ctt	gct	gct	gca	aca	aat	gtt	gct	gct	ggc	ctg	ttt	aag	aaa	tta	384	
Ile	Leu	Ala	Ala	Ala	Thr	Asn	Val	Ala	Ala	Gly	Leu	Phe	Lys	Lys	Leu		
		115					120					125					
ttg	gtt	ggt	tct	ctt	gtt	atg	ctt	gtg	ttt	ggt	tac	atg	ggt	gag	gca	432	
Leu	Val	Gly	Ser	Leu	Val	Met	Leu	Val	Phe	Gly	Tyr	Met	Gly	Glu	Ala		
	130					135					140						
gga	att	atg	aac	gct	tgg	cct	gca	ttc	att	att	ggg	tgt	tta	gct	tgg	480	
Gly	Ile	Met	Asn	Ala	Trp	Pro	Ala	Phe	Ile	Ile	Gly	Cys	Leu	Ala	Trp		
145					150					155					160		
gta	tac	atg	att	tat	gaa	cta	tat	gct	gga	gaa	gga	aaa	tct	gca	tgt	528	
Val	Tyr	Met	Ile	Tyr	Glu	Leu	Tyr	Ala	Gly	Glu	Gly	Lys	Ser	Ala	Cys		
				165					170					175			
aat	act	gca	agt	cct	tcg	gtt	caa	tca	gct	tac	aac	aca	atg	atg	gct	576	
Asn	Thr	Ala	Ser	Pro	Ser	Val	Gln	Ser	Ala	Tyr	Asn	Thr	Met	Met	Ala		
			180					185					190				
atc	ata	gtc	ttc	ggt	tgg	gca	att	tat	cct	gta	ggt	tat	ttc	aca	ggt	624	
Ile	Ile	Val	Phe	Gly	Trp	Ala	Ile	Tyr	Pro	Val	Gly	Tyr	Phe	Thr	Gly		
		195					200					205					
tac	cta	atg	ggt	gac	ggt	gga	tca	gct	ctt	aac	tta	aac	ctt	att	tat	672	
Tyr	Leu	Met	Gly	Asp	Gly	Gly	Ser	Ala	Leu	Asn	Leu	Asn	Leu	Ile	Tyr		
	210					215					220						
aac	ctt	gct	gac	ttt	gtt	aac	aag	att	cta	ttt	ggt	tta	att	ata	tgg	720	
Asn	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile	Trp		
225					230					235					240		
aat	gtt	gct	gtt	aaa	gaa	tct	tct	aat	gct							750	
Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala								
				245					250								

<210> 25
 <211> 250
 <212> PRT
 <213> Naturally occurring gamma proteobacterium

<400> 25

Met. Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
 100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu
 115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
 130 135 140

Gly Ile Met Asn Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
 145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys
 165 170 175

Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala

180

185

190

Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
 210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 26
 <211> 750
 <212> DNA
 <213> Naturally occurring gamma proteobacterium

<220>
 <221> CDS
 <222> (1)..(750)
 <223> Proteorhodopsin variant from pcr clone MB0m2

<400> 26
 atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca 48
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15
 ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt 96
 Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30
 tct ttt tgg tta gtt act gct gct tta tta gca tct act gta ttt ttc 144
 Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45
 ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192
 Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60
 gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg 240
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80
 aga ggg gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac 288

Arg	Gly	Val	Trp	Ile	Glu	Thr	Gly	Asp	Ser	Pro	Thr	Val	Phe	Arg	Tyr		
				85					90					95			
att	gat	tgg	tta	cta	aca	gtt	cct	cta	tta	ata	tgt	gaa	ttc	tac	tta		336
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Leu	Ile	Cys	Glu	Phe	Tyr	Leu		
			100					105					110				
att	ctt	gct	gct	gct	act	aat	gtt	gct	gct	ggc	ctg	ttt	aag	aaa	tta		384
Ile	Leu	Ala	Ala	Ala	Thr	Asn	Val	Ala	Ala	Gly	Leu	Phe	Lys	Lys	Leu		
		115					120					125					
ttg	gtt	ggg	tct	ctt	gtt	atg	ctt	gtg	ttt	ggg	tac	atg	ggg	gaa	gca		432
Leu	Val	Gly	Ser	Leu	Val	Met	Leu	Val	Phe	Gly	Tyr	Met	Gly	Glu	Ala		
	130					135					140						
gga	att	atg	aac	gct	tgg	ggg	gca	ttc	gtt	att	ggg	tgt	tta	gct	tgg		480
Gly	Ile	Met	Asn	Ala	Trp	Gly	Ala	Phe	Val	Ile	Gly	Cys	Leu	Ala	Trp		
145					150					155					160		
gta	tac	atg	att	tat	gag	ctt	tgg	ctt	gga	gaa	gga	aaa	gct	gcg	tgt		528
Val	Tyr	Met	Ile	Tyr	Glu	Leu	Trp	Leu	Gly	Glu	Gly	Lys	Ala	Ala	Cys		
				165					170					175			
aat	aca	gca	agt	cct	gct	gtt	cag	tca	gct	tac	aac	aca	atg	atg	atg		576
Asn	Thr	Ala	Ser	Pro	Ala	Val	Gln	Ser	Ala	Tyr	Asn	Thr	Met	Met	Met		
			180					185						190			
atc	atc	atc	ttt	ggg	tgg	gca	att	tat	cct	gta	ggg	tat	ttc	aca	ggg		624
Ile	Ile	Ile	Phe	Gly	Trp	Ala	Ile	Tyr	Pro	Val	Gly	Tyr	Phe	Thr	Gly		
		195					200					205					
tac	cta	atg	ggg	gac	ggg	gga	tca	gca	ctt	aac	tta	aac	ctt	atc	tat		672
Tyr	Leu	Met	Gly	Asp	Gly	Gly	Ser	Ala	Leu	Asn	Leu	Asn	Leu	Ile	Tyr		
	210					215					220						
aac	ctt	gct	gac	ttt	gtt	aac	aag	att	cta	ttt	ggg	tta	att	ata	tgg		720
Asn	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile	Trp		
225					230					235					240		
aat	gtt	gct	gtt	aaa	gaa	tct	tct	aat	gct								750
Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala								
				245					250								

<210> 27

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 27

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr

1	5	10	15												
Phe	Ala	Ala	Gly	Gly	Gly	Asp	Leu	Asp	Ala	Ser	Asp	Tyr	Thr	Gly	Val
			20					25						30	
Ser	Phe	Trp	Leu	Val	Thr	Ala	Ala	Leu	Leu	Ala	Ser	Thr	Val	Phe	Phe
		35					40					45			
Phe	Val	Glu	Arg	Asp	Arg	Val	Ser	Ala	Lys	Trp	Lys	Thr	Ser	Leu	Thr
	50					55					60				
Val	Ser	Gly	Leu	Val	Thr	Gly	Ile	Ala	Phe	Trp	His	Tyr	Met	Tyr	Met
65					70					75					80
Arg	Gly	Val	Trp	Ile	Glu	Thr	Gly	Asp	Ser	Pro	Thr	Val	Phe	Arg	Tyr
				85					90					95	
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Leu	Ile	Cys	Glu	Phe	Tyr	Leu
			100					105						110	
Ile	Leu	Ala	Ala	Ala	Thr	Asn	Val	Ala	Ala	Gly	Leu	Phe	Lys	Lys	Leu
		115					120						125		
Leu	Val	Gly	Ser	Leu	Val	Met	Leu	Val	Phe	Gly	Tyr	Met	Gly	Glu	Ala
	130					135					140				
Gly	Ile	Met	Asn	Ala	Trp	Gly	Ala	Phe	Val	Ile	Gly	Cys	Leu	Ala	Trp
145					150					155					160
Val	Tyr	Met	Ile	Tyr	Glu	Leu	Trp	Leu	Gly	Glu	Gly	Lys	Ala	Ala	Cys
				165					170					175	
Asn	Thr	Ala	Ser	Pro	Ala	Val	Gln	Ser	Ala	Tyr	Asn	Thr	Met	Met	Met
			180					185					190		
Ile	Ile	Ile	Phe	Gly	Trp	Ala	Ile	Tyr	Pro	Val	Gly	Tyr	Phe	Thr	Gly
		195					200					205			
Tyr	Leu	Met	Gly	Asp	Gly	Gly	Ser	Ala	Leu	Asn	Leu	Asn	Leu	Ile	Tyr

210

215

220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 28

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> Proteorhodopsin variant from pcr clone MB20m2; GenBank #AF349985

<400> 28

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca 48
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt 96
 Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30

tct ttt tgg tta gtt act gct gct tta tta gca tct act gta ttt ttc 144
 Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45

ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192
 Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg 240
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80

aga ggg gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac 288
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95

att gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
 100 105 110

att ctt gct gct gca act aat gtt gct gct ggc ctg ttt aag aaa tta 384
 Ile Leu Ala Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu
 115 120 125

ttg gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gag gca 432
 Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
 130 135 140

gga att atg aac gct tgg ggt gca ttc gtt att ggg tgt tta gct tgg 480
 Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp
 145 150 155 160

gta tac atg att tat gaa cta tgg gct gga gaa ggc aag gct gca tgt 528
 Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ala Ala Cys
 165 170 175

aat act gca agt cct gct gtg caa tca gct tac aac aca atg atg tat 576
 Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
 180 185 190

ata atc atc ttt ggt tgg gca att tat cct gta ggt tat ttc aca ggt 624
 Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205

tac cta atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat 672
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
 210 215 220

aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720
 Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240

aat gtt gct gtt aaa gaa tct tct aat gct 750
 Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 29

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 29

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu
115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp
145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ala Ala Cys
165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
180 185 190

Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 30
<211> 750
<212> DNA
<213> Naturally occurring gamma proteobacterium

<220>
<221> CDS
<222> (1)..(750)
<223> Proteorhodopsin variant from pcr clone MB20m5; GenBank#AF349986

<400> 30
atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca 48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1 5 10 15

ttt gct gca ggt ggc ggt gac ctt gat gct agt gat tac act ggt gtt 96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20 25 30

tct ttt tgg tta gtt aca gct gct cta tta gca tct act gta ttt ttc 144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45

ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg 240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

aga ggg gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac 288
Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85 90 95

att gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta 336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

att ctt gct gct gct act aat gtt gct gga tca tta ttt aag aaa tta 384
Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
115 120 125

cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca 432
Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

caa att atg gct gca tgg cct gca ttc att att ggg tgt tta gct tgg 480
 Gln Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
 145 150 155 160
 gta tac atg att tat gaa cta tat gct gga gaa gga aaa tct gca tgt 528
 Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys
 165 170 175
 aat act gca agt cct tcg gtt caa tca gct tac aac aca atg atg gct 576
 Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala
 180 185 190
 atc ata gtc ttc ggt tgg gca att tat cct gta ggt tat ttc aca ggt 624
 Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205
 tac cta atg ggt gac ggt ggg tca gct ctt aac tta aac ctt att tat 672
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
 210 215 220
 aac ctt gct gac ttt gtt aac aag att cta ctt ggt tta att ata tgg 720
 Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Leu Gly Leu Ile Ile Trp
 225 230 235 240
 aat gtt gct gtt aaa gaa tct tct aat gct 750
 Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 31

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 31

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

Gln Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys
165 170 175

Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala
180 185 190

Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Leu Gly Leu Ile Ile Trp
225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 32

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> Proteorhodopsin variant from pcr clone MB20m12; GenBank #AF349987

<400> 32

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca	48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr	
1 5 10 15	
ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt	96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val	
20 25 30	
tct ttt tgg tta gtt act gct gct tta tta gca tct act gta ttt ttc	144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe	
35 40 45	
ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act	192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	
gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg	240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met	
65 70 75 80	
aga ggg gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac	288
Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr	
85 90 95	
att gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta	336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu	
100 105 110	
att ctt gct gct gca gct aat gtt gct gga tca tta ttt aag aaa tta	384
Ile Leu Ala Ala Ala Ala Asn Val Ala Gly Ser Leu Phe Lys Lys Leu	
115 120 125	
cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca	432
Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala	
130 135 140	
gga atc atg gct gca tgg cct gca ttc att att ggg tgt tta gct tgg	480
Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp	
145 150 155 160	
gta tac atg att tat gaa tta tgg gct gga gaa gga aaa tct gca tgt	528
Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys	
165 170 175	

aat act gca agt cct gct gtg caa tca gcc tac aac aca atg atg tat 576
 Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
 180 185 190

att atc atc ttt ggt tgg gcg att tat cct gta ggt tat ttc aca ggt 624
 Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205

tac ttg atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat 672
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
 210 215 220

aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720
 Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240

aat gtt gct gtt aaa gaa tct tct aat gct 750
 Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 33

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 33

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Ala Ala Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys
165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
180 185 190

Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 34

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> Proteorhodopsin variant from pcr clone MB40m1; GenBank #AF349988

<400> 34

atg ggt aaa tta tta ctg ata ata ggt agt gtt att gca ctt cct aca	48
Met Gly Lys Leu Leu Leu Ile Ile Gly Ser Val Ile Ala Leu Pro Thr	
1 5 10 15	
ttt gct gca ggt ggc ggt gac ctt gat gct agt gat tac act ggt gtt	96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val	
20 25 30	
tct ttt tgg tta gtt aca gct gct cta tta gca tct act gta ttt ttc	144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe	
35 40 45	
ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act	192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	
gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg	240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met	
65 70 75 80	
aga gga gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac	288
Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr	
85 90 95	
att gat tgg tta cta aca gtt cct tta tta ata tgt gaa ttc tac tta	336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu	
100 105 110	
att ctt gct gct gca act aat gtt gcc ggc tca tta ttt aag aaa ctt	384
Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu	
115 120 125	
cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca	432
Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala	
130 135 140	
gga att atg gca gct tgg cct gca ttc att att ggg tgt tta gct tgg	480
Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp	
145 150 155 160	
gta tat atg att tat gaa cta tat gct gga gaa gga aaa tct gca tgt	528
Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys	
165 170 175	
aat aca gca agt cct gct gtg caa tca gct tac aac aca atg atg tat	576
Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr	
180 185 190	
att atc gtc ttt ggt tgg gcg att tat cct gta ggt tat ttc aca ggt	624
Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly	
195 200 205	

tac ctg atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat 672
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
 210 215 220

aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720
 Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240

aat gtt gct gtt aaa gaa tct tct aat gct 750
 Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 35

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 35

Met Gly Lys Leu Leu Leu Ile Ile Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
 100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
 115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
 130 135 140

Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
 145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys
 165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
 180 185 190

Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
 210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 36

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> Proteorhodopsin variant from pcr clone MB40m5;p GenBank #AF349989

<400> 36

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca 48
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt 96
 Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30

tct ttt tgg tta gtt act gct gct cta tta gca tct act gta ttt ttc	144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe	
35 40 45	
ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act	192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	
gta tcg ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg	240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met	
65 70 75 80	
aga ggg gta tgg att gag act ggt gat tcg cca act gta ttt aga tac	288
Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr	
85 90 95	
att gat tgg tta cta aca gtt cct cta ttg ata tgt gaa ttc tac tta	336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu	
100 105 110	
att ctt gct gct gca aca aat gtt gct gct ggc ctg ttt aag aaa tta	384
Ile Leu Ala Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu	
115 120 125	
ttg gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gag gca	432
Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala	
130 135 140	
gga att atg aac gct tgg ggt gca ttc gtt att ggg tgt tta gct tgg	480
Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp	
145 150 155 160	
gta tac atg att tat gaa cta tgg gct gga gaa ggc aag gct gca tgt	528
Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ala Ala Cys	
165 170 175	
aat act gca agt cct gct gtg caa tca gct tac aac aca atg atg tat	576
Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr	
180 185 190	
ata atc atc ttt ggt tgg gca att tat cct gta ggt tat ttc aca ggt	624
Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly	
195 200 205	
tac cta atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat	672
Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr	
210 215 220	
aac ctt gct gac ttt gtt aac aag aat cta ttt ggt tta att ata tgg	720
Asn Leu Ala Asp Phe Val Asn Lys Asn Leu Phe Gly Leu Ile Ile Trp	
225 230 235 240	

aat gtt gct gtt aaa gaa tct tct aat gct
 Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

750

<210> 37
 <211> 250
 <212> PRT
 <213> Naturally occurring gamma proteobacterium
 <400> 37

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
 100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu
 115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
 130 135 140

Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp
 145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ala Ala Cys
 165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
 180 185 190

Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
 210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Asn Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 38

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> Proteorhodopsin variant from pcr clone MB40m12; GenBank # AF34999

<400> 38

atg ggt aaa tta tta cgg ata tta ggt agt gtt att gca ctt cct aca 48
 Met Gly Lys Leu Leu Arg Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

ttt gct gca ggt ggc ggt gac ctt gat gct agt gat tac act ggt gtt 96
 Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30

tct ttt tgg tta gtt aca gct gct cta tta gca tct act gta ttt ttc 144
 Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45

ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192
 Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr

50	55	60	
gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tat atg			240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met			
65	70	75	80
aga gga gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac			288
Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr			
	85	90	95
att gat tgg tta cta aca gtt cct tta tta ata tgt gaa ttc tac tta			336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu			
	100	105	110
att ctt gct gct gca act aat gtt gct gga tca tta ttt aag aaa tta			384
Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu			
	115	120	125
cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca			432
Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala			
	130	135	140
gga atc atg gct gca tgg cct gca ttc att att ggg tgt tta gct tgg			480
Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp			
	145	150	155
gta tac atg att tat gaa cta tgg gct gga gaa gga aaa tct gca tgt			528
Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys			
	165	170	175
aat act gca agt cct gct gtg caa tca gct tac aac aca atg atg tat			576
Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr			
	180	185	190
atc atc atc gtt ggt tgg gcg att tat cct gta ggt tat ttc aca ggt			624
Ile Ile Ile Val Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly			
	195	200	205
tac ctg atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat			672
Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr			
	210	215	220
aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg			720
Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp			
	225	230	235
aat gtt gct gtt aaa gaa tct tct aat gct			750
Asn Val Ala Val Lys Glu Ser Ser Asn Ala			
	245	250	

<210> 39

<211> 250
 <212> PRT
 <213> Naturally occurring gamma proteobacterium
 <400> 39

Met Gly Lys Leu Leu Arg Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
 100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
 115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
 130 135 140

Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
 145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys
 165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
 180 185 190

Ile Ile Ile Val Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 40

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> Proteorhodopsin variant from pcr clone MB100m5; GenBank #AF349991

<400> 40

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca 48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1 5 10 15

ttt gct gca ggt ggc ggt gac ctt gat gct agt gat tac act ggt gtt 96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20 25 30

tct ttt tgg tta gtt aca gct gct cta tta gca tct act gta ttt ttc 144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45

ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg 240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

aga gga gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac 288
Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr

85										90					95					
att	gat	tgg	tta	cta	aca	gtt	cct	tta	tta	ata	tgt	gaa	ttc	tac	tta	336				
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Leu	Ile	Cys	Glu	Phe	Tyr	Leu					
			100							105				110						
att	ctt	gct	gct	gca	act	aat	gtt	gcc	ggc	tca	tta	ttt	aag	aaa	ctt	384				
Ile	Leu	Ala	Ala	Ala	Thr	Asn	Val	Ala	Gly	Ser	Leu	Phe	Lys	Lys	Leu					
			115							120				125						
cta	gtt	ggt	tct	ctt	gtt	atg	ctt	gtg	ttt	ggt	tac	atg	ggt	gaa	gca	432				
Leu	Val	Gly	Ser	Leu	Val	Met	Leu	Val	Phe	Gly	Tyr	Met	Gly	Glu	Ala					
			130							135				140						
gga	att	atg	gca	gct	tgg	cct	gca	ttc	att	att	ggg	tgt	tta	gct	tgg	480				
Gly	Ile	Met	Ala	Ala	Trp	Pro	Ala	Phe	Ile	Ile	Gly	Cys	Leu	Ala	Trp					
			145							150				155						
gta	tac	atg	att	tat	gaa	cta	tat	gct	gga	gaa	gga	aaa	tct	gca	tgt	528				
Val	Tyr	Met	Ile	Tyr	Glu	Leu	Tyr	Ala	Gly	Glu	Gly	Lys	Ser	Ala	Cys					
			165							170				175						
aat	act	gca	agt	cct	tcg	gtt	caa	tca	gct	tac	aac	aca	atg	atg	gct	576				
Asn	Thr	Ala	Ser	Pro	Ser	Val	Gln	Ser	Ala	Tyr	Asn	Thr	Met	Met	Ala					
			180							185				190						
atc	ata	gtc	ttc	ggt	tgg	gca	att	tat	cct	gta	ggt	tat	ttc	aca	ggt	624				
Ile	Ile	Val	Phe	Gly	Trp	Ala	Ile	Tyr	Pro	Val	Gly	Tyr	Phe	Thr	Gly					
			195							200				205						
tac	cta	atg	ggt	gac	ggt	gga	tca	gct	ctt	aac	tta	aac	ctt	att	tat	672				
Tyr	Leu	Met	Gly	Asp	Gly	Gly	Ser	Ala	Leu	Asn	Leu	Asn	Leu	Ile	Tyr					
			210							215				220						
aac	ctt	gct	gac	ttt	gtt	aac	aag	att	cta	ttt	ggt	tta	att	ata	tgg	720				
Asn	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile	Trp					
			225							230				235						
aat	gtt	gct	gtt	aaa	gaa	tct	tct	aat	gct							750				
Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala											
			245							250										

<210> 41

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 41

Met	Gly	Lys	Leu	Leu	Leu	Ile	Leu	Gly	Ser	Val	Ile	Ala	Leu	Pro	Thr
1				5					10					15	

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys
165 170 175

Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala
180 185 190

Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 42

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> Proteorhodopsin variant from pcr clone MB100m7; GenBank #AF349992

<400> 42

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca 48
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt 96
 Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30

tct ttt tgg tta gtt act gct gct tta tta gca tct act gta ttt ttc 144
 Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45

ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192
 Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg 240
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80

aga ggg gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac 288
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95

att gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
 100 105 110

att ctt gct gct gct act aat gtt gcc ggc tca tta ttt aag aaa ctt 384
 Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu

115	120	125	
cta gtt ggt tct ctt gtt atg	ctt gtg ttt ggt tac atg ggt gaa gca		432
Leu Val Gly Ser Leu Val Met	Leu Val Phe Gly Tyr Met Gly Glu Ala		
130	135	140	
gga att atg gca gct tgg cct gca ttc att att	ggg tgt tta gct tgg		480
Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile	Gly Cys Leu Ala Trp		
145	150	155	160
gta tac atg att tat gaa cta tat gct gga gaa gga aaa tct gca tgt			528
Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys			
165	170	175	
aat act gca agt cct tcg gtt caa tca gct tac aac aca atg atg gct			576
Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala			
180	185	190	
atc ata gtc ttc ggt tgg gca att tat cct gta ggt tat ttc aca ggt			624
Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly			
195	200	205	
tac cta atg ggt gac ggt gga tca gct ctt aac tta aac ctt att tat			672
Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr			
210	215	220	
aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg			720
Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp			
225	230	235	240
aat gct gct gtt aaa gaa tct tct aat gct			750
Asn Ala Ala Val Lys Glu Ser Ser Asn Ala			
245	250		

<210> 43

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 43

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys
165 170 175

Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala
180 185 190

Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
225 230 235 240

Asn Ala Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 44
 <211> 750
 <212> DNA
 <213> Naturally occurring gamma proteobacterium

 <220>
 <221> CDS
 <222> (1)..(750)
 <223> Proteorhodopsin variant from pcr clone MB100m9; GenBank #AF349993

 <400> 44
 atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca 48
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

 ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt 96
 Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30

 tct ttt tgg tta gtt act gct gct tta tta gca tct act gta ttt ttc 144
 Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45

 ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192
 Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

 gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tat atg 240
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80

 aga ggg gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac 288
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95

 ata gat tgg tta cta aca gtt cct tta tta ata tgt gaa ttc tac tta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
 100 105 110

 att ctt gcc gct gca act aat gtt gct gga tca tta ttt aag aaa tta 384
 Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
 115 120 125

 ctt gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca 432
 Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
 130 135 140

 gga atc atg gct gca tgg cct gca ttc att att ggg tgt tta gct tgg 480
 Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp

145	150	155	160	
gta tac atg att tat gaa cta tgg gct gga gaa gga aaa tct gca tgt				528
Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys	165	170	175	
aat act gca agt cct gct gtg caa tca gct tac aac aca atg atg tat				576
Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr	180	185	190	
atc atc atc ttt ggt tgg gcg att tat cct gta ggt tat ttc aca ggt				624
Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly	195	200	205	
tac ctt atg ggt gac ggt gga tca gca ctt aac tta aac ctt att tat				672
Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr	210	215	220	
aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg				720
Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp	225	230	235	240
aat gtt gct gtt aaa gaa tct tct aat gct				750
Asn Val Ala Val Lys Glu Ser Ser Asn Ala	245	250		

<210> 45
 <211> 250
 <212> PRT
 <213> Naturally occurring gamma proteobacterium

<400> 45

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130 135 140

Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys
165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
180 185 190

Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 46

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> Proteorhodopsin variant from pcr clone MB100m10; GenBank #AF34999

<400> 46

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca	48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr	
1 5 10 15	

ttt gct gca ggt ggc ggt gac ctt gat gct agt gat tac act ggt gtt	96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val	
20 25 30	

tct ttt tgg tta gtt aca gct gct cta tta gcg tct act gta ttt ttc	144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe	
35 40 45	

ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act	192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	

gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tat atg	240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met	
65 70 75 80	

aga gga gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac	288
Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr	
85 90 95	

att gat tgg tta cta aca gtt cct tta tta ata tgt gaa ttc tac tta	336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu	
100 105 110	

att ctt gct gct gca act aat gtt gcc ggc tca tta ttt aag aaa ctt	384
Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu	
115 120 125	

cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca	432
Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala	
130 135 140	

gga ata atg gcg gct tgg cct gca ttc atc gtt gga tgt tta gca tgg	480
Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Val Gly Cys Leu Ala Trp	
145 150 155 160	

gta tat atg att tat gaa cta tgg gct ggt gaa gga aaa tct gca tgt	528
Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys	
165 170 175	

aat act gca agt cct gct gta cag tca gct tac aac aca atg atg tat	576
---	-----

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
 180 185 190
 atc atc atc gtt ggt tgg gca att tat cct gta ggt tat ttc aca ggt 624
 Ile Ile Ile Val Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205
 tac cta atg ggt gac ggt gga tca gct ctt aat cta aac ctt att tat 672
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
 210 215 220
 aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720
 Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240
 aat gtt gct gtt aaa gaa tct tct aat gct 750
 Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 47
 <211> 250
 <212> PRT
 <213> Naturally occurring gamma proteobacterium
 <400> 47

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
 1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
 65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
 85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu

100

105

110

Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
 115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
 130 135 140

Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Val Gly Cys Leu Ala Trp
 145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys
 165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
 180 185 190

Ile Ile Ile Val Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
 195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
 210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
 225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 48

<211> 753

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone PALB1; GenBank #AF349995

<400> 48

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca

48

Met	Gly	Lys	Leu	Leu	Leu	Ile	Leu	Gly	Ser	Ala	Ile	Ala	Leu	Pro	Ser		
1				5					10					15			
ttt	gct	gct	gct	ggt	ggc	gat	cta	gat	ata	agt	gat	act	ggt	ggt	ggt		96
Phe	Ala	Ala	Ala	Gly	Gly	Asp	Leu	Asp	Ile	Ser	Asp	Thr	Val	Gly	Val		
			20					25					30				
tca	ttc	tgg	ctg	ggt	aca	gct	ggt	atg	tta	gcg	gca	act	gta	ttc	ttt		144
Ser	Phe	Trp	Leu	Val	Thr	Ala	Gly	Met	Leu	Ala	Ala	Thr	Val	Phe	Phe		
		35					40					45					
ttt	gta	gaa	aga	gac	caa	gtc	agc	gct	aag	tgg	aaa	act	tca	ctt	act		192
Phe	Val	Glu	Arg	Asp	Gln	Val	Ser	Ala	Lys	Trp	Lys	Thr	Ser	Leu	Thr		
	50					55					60						
gta	tct	ggt	tta	att	act	ggt	ata	gct	ttt	tgg	cat	tat	ctc	tac	atg		240
Val	Ser	Gly	Leu	Ile	Thr	Gly	Ile	Ala	Phe	Trp	His	Tyr	Leu	Tyr	Met		
65					70				75						80		
aga	ggt	ggt	tgg	ata	gat	act	ggt	gat	aca	cca	aca	gta	ttt	aga	tat		288
Arg	Gly	Val	Trp	Ile	Asp	Thr	Gly	Asp	Thr	Pro	Thr	Val	Phe	Arg	Tyr		
				85				90						95			
att	gat	tgg	cta	tta	act	ggt	cca	tta	caa	atg	ggt	gag	ttc	tat	cta		336
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Gln	Met	Val	Glu	Phe	Tyr	Leu		
			100					105					110				
att	ctt	gct	gct	tgt	aca	agt	ggt	gct	gct	tca	tta	ttt	aag	aag	ctt		384
Ile	Leu	Ala	Ala	Cys	Thr	Ser	Val	Ala	Ala	Ser	Leu	Phe	Lys	Lys	Leu		
		115					120					125					
cta	gct	ggt	tca	tta	gta	atg	tta	ggt	gct	gga	ttt	gca	ggc	gaa	gct		432
Leu	Ala	Gly	Ser	Leu	Val	Met	Leu	Gly	Ala	Gly	Phe	Ala	Gly	Glu	Ala		
	130					135					140						
ggt	tta	gct	cct	gta	tta	cct	gct	ttc	att	ctt	ggt	atg	gct	ggt	tgg		480
Gly	Leu	Ala	Pro	Val	Leu	Pro	Ala	Phe	Ile	Leu	Gly	Met	Ala	Gly	Trp		
145					150				155						160		
tta	tac	atg	att	tat	gag	cta	cat	atg	ggt	gaa	ggt	aag	gct	gct	gta		528
Leu	Tyr	Met	Ile	Tyr	Glu	Leu	His	Met	Gly	Glu	Gly	Lys	Ala	Ala	Val		
			165					170					175				
agt	act	gca	agt	cct	gct	ggt	aac	tct	gct	tac	aat	gca	atg	atg	aag		576
Ser	Thr	Ala	Ser	Pro	Ala	Val	Asn	Ser	Ala	Tyr	Asn	Ala	Met	Met	Lys		
			180					185					190				
att	att	ggt	att	gga	tgg	gca	att	tat	cct	gct	gga	tat	gct	gct	ggt		624
Ile	Ile	Val	Ile	Gly	Trp	Ala	Ile	Tyr	Pro	Ala	Gly	Tyr	Ala	Ala	Gly		
		195					200					205					
tac	cta	atg	agt	ggt	gac	ggt	gta	tac	gct	tca	aac	tta	aac	ctt	ata		672

Tyr Leu Met Ser Gly Asp Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
 210 215 220
 tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att 720
 Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
 225 230 235 240
 tgg aat gtt gct gtt aaa gaa tct tct aat gct 753
 Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 49
 <211> 251
 <212> PRT
 <213> Naturally occurring gamma proteobacterium
 <400> 49

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
 100 105 110

Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
 115 120 125

Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala

130

135

140

Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Leu Gly Met Ala Gly Trp
 145 150 155 160

Leu Tyr Met Ile Tyr Glu Leu His Met Gly Glu Gly Lys Ala Ala Val
 165 170 175

Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Lys
 180 185 190

Ile Ile Val Ile Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
 195 200 205

Tyr Leu Met Ser Gly Asp Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
 210 215 220

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
 225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 50

<211> 753

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone PALB2; GenBank #AF349996

<400> 50

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15

ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
 Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30

tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gtg ttc ttt 144

Ser	Phe	Trp	Leu	Val	Thr	Ala	Gly	Met	Leu	Ala	Ala	Thr	Val	Phe	Phe	
		35					40					45				
ttt	gta	gaa	aga	gac	caa	gtc	agc	gct	gag	tgg	aaa	act	tca	ctt	act	192
Phe	Val	Glu	Arg	Asp	Gln	Val	Ser	Ala	Glu	Trp	Lys	Thr	Ser	Leu	Thr	
	50					55					60					
gta	tct	ggg	tta	att	act	ggg	ata	gct	ttt	tgg	cat	tat	ctc	tat	atg	240
Val	Ser	Gly	Leu	Ile	Thr	Gly	Ile	Ala	Phe	Trp	His	Tyr	Leu	Tyr	Met	
65					70					75					80	
aga	ggg	gtt	tgg	ata	gat	act	ggg	gat	acc	cca	aca	gta	ttc	aga	tat	288
Arg	Gly	Val	Trp	Ile	Asp	Thr	Gly	Asp	Thr	Pro	Thr	Val	Phe	Arg	Tyr	
				85					90					95		
att	gat	tgg	tta	tta	act	gtt	cca	tta	caa	atg	gtt	gag	ttc	tat	cta	336
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Gln	Met	Val	Glu	Phe	Tyr	Leu	
			100					105					110			
att	ctt	gct	gct	tgt	aca	agt	gtt	gct	gct	tca	tta	ttt	aag	aag	ctt	384
Ile	Leu	Ala	Ala	Cys	Thr	Ser	Val	Ala	Ala	Ser	Leu	Phe	Lys	Lys	Leu	
		115					120					125				
cta	gct	ggg	tca	tta	gta	atg	tta	ggg	gct	gga	ttt	gca	ggc	gaa	gct	432
Leu	Ala	Gly	Ser	Leu	Val	Met	Leu	Gly	Ala	Gly	Phe	Ala	Gly	Glu	Ala	
	130					135					140					
gga	tta	gct	cct	gta	tta	cct	gct	ttc	att	att	ggg	atg	gct	gga	tgg	480
Gly	Leu	Ala	Pro	Val	Leu	Pro	Ala	Phe	Ile	Ile	Gly	Met	Ala	Gly	Trp	
145					150				155					160		
tta	tac	atg	att	tat	gag	cta	tat	atg	ggg	gaa	ggg	aag	gct	gct	gta	528
Leu	Tyr	Met	Ile	Tyr	Glu	Leu	Tyr	Met	Gly	Glu	Gly	Lys	Ala	Ala	Val	
				165					170					175		
agt	act	gca	agt	cct	gct	gtt	aac	tct	gca	tac	aac	gca	atg	atg	atg	576
Ser	Thr	Ala	Ser	Pro	Ala	Val	Asn	Ser	Ala	Tyr	Asn	Ala	Met	Met	Met	
			180					185					190			
att	att	gtt	gtt	gga	tgg	gca	att	tat	cct	gct	gga	tat	gct	gct	ggg	624
Ile	Ile	Val	Val	Gly	Trp	Ala	Ile	Tyr	Pro	Ala	Gly	Tyr	Ala	Ala	Gly	
		195					200					205				
tac	cta	atg	ggg	ggc	gaa	ggg	gta	tac	gct	tca	aac	tta	aac	ctt	ata	672
Tyr	Leu	Met	Gly	Gly	Glu	Gly	Val	Tyr	Ala	Ser	Asn	Leu	Asn	Leu	Ile	
	210					215					220					
tat	aac	ctt	gct	gac	ttt	gtt	aac	aag	att	cta	ttt	ggg	ttg	atc	att	720
Tyr	Asn	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile	
225					230				235					240		
tgg	aat	gtt	gct	gtt	aaa	gaa	tct	tct	aat	gct						753

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 51

<211> 251

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 51

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Glu Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
115 120 125

Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
130 135 140

Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
145 150 155 160

Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val

165

170

175

Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met
 180 185 190

Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
 195 200 205

Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
 210 215 220

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
 225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 52

<211> 753

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone PALB5; GenBank#AF349997

<400> 52

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15

ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
 Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30

tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gtg ttc ttt 144
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
 35 40 45

ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act 192
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

gta tct ggt tta att act ggt ata gcc ttt tgg cat tat ctc tat atg 240

Val	Ser	Gly	Leu	Ile	Thr	Gly	Ile	Ala	Phe	Trp	His	Tyr	Leu	Tyr	Met	
65					70				75						80	
aga	ggt	ggt	tgg	ata	gac	act	ggt	gat	acc	cca	aca	gta	ttc	aga	tat	288
Arg	Gly	Val	Trp	Ile	Asp	Thr	Gly	Asp	Thr	Pro	Thr	Val	Phe	Arg	Tyr	
				85					90					95		
att	gat	tgg	tta	tta	act	ggt	cca	tta	caa	atg	ggt	gag	ttc	tat	cta	336
Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Gln	Met	Val	Glu	Phe	Tyr	Leu	
			100					105					110			
att	ctt	gct	gct	tgt	aca	aat	ggt	gct	gct	tca	tta	ttt	aag	aag	ctt	384
Ile	Leu	Ala	Ala	Cys	Thr	Asn	Val	Ala	Ala	Ser	Leu	Phe	Lys	Lys	Leu	
		115					120					125				
cta	gct	ggt	tca	tta	gta	atg	tta	ggt	gct	gga	ttt	gca	ggc	gaa	gct	432
Leu	Ala	Gly	Ser	Leu	Val	Met	Leu	Gly	Ala	Gly	Phe	Ala	Gly	Glu	Ala	
	130					135					140					
gga	tta	gct	cct	gta	tgg	cct	gct	ttc	att	att	ggt	atg	gct	gga	tgg	480
Gly	Leu	Ala	Pro	Val	Trp	Pro	Ala	Phe	Ile	Ile	Gly	Met	Ala	Gly	Trp	
145					150					155					160	
tta	tac	atg	att	tat	gag	cta	tat	atg	ggt	gaa	ggt	aag	gct	gct	gta	528
Leu	Tyr	Met	Ile	Tyr	Glu	Leu	Tyr	Met	Gly	Glu	Gly	Lys	Ala	Ala	Val	
				165					170					175		
agt	act	gca	agt	cct	gct	ggt	aac	tct	gca	tac	aac	gca	atg	atg	atg	576
Ser	Thr	Ala	Ser	Pro	Ala	Val	Asn	Ser	Ala	Tyr	Asn	Ala	Met	Met	Met	
			180					185					190			
att	att	ggt	ggt	gga	tgg	gca	att	tat	cct	gct	gga	tat	gct	gct	ggt	624
Ile	Ile	Val	Val	Gly	Trp	Ala	Ile	Tyr	Pro	Ala	Gly	Tyr	Ala	Ala	Gly	
		195				200						205				
tac	cta	atg	ggt	ggc	gaa	ggt	gta	tac	gct	tca	aac	cta	aac	ctt	ata	672
Tyr	Leu	Met	Gly	Gly	Glu	Gly	Val	Tyr	Ala	Ser	Asn	Leu	Asn	Leu	Ile	
	210					215					220					
tat	aac	ctt	gct	gac	ttt	ggt	aac	aag	att	cta	ttt	ggt	ttg	atc	att	720
Tyr	Asn	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile	
225					230					235					240	
tgg	aat	ggt	gct	ggt	aaa	gaa	tct	tct	aat	gct						753
Trp	Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala						
				245					250							

<210> 53

<211> 251

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 53

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
 100 105 110

Ile Leu Ala Ala Cys Thr Asn Val Ala Ala Ser Leu Phe Lys Lys Leu
 115 120 125

Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
 130 135 140

Gly Leu Ala Pro Val Trp Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
 145 150 155 160

Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
 165 170 175

Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met
 180 185 190

Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly

195	200	205
Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile		
210	215	220
Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile		
225	230	235 240
Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala		
245	250	

<210> 54
 <211> 753
 <212> DNA
 <213> Naturally occurring gamma proteobacterium

 <220>
 <221> CDS
 <222> (1)..(753)
 <223> Proteorhodopsin variant from pcr clone PalB7; GenBank #AF349999

<400> 54	
atg ggt aaa tta tta ctg ata tta ggt agt gct att gcg ctt cca tca	48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser	
1 5 10 15	
ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt	96
Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val	
20 25 30	
tca ttc tgg ctg gtt acg gct ggt atg tta gcg gca act gta ttc ttt	144
Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe	
35 40 45	
ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act	192
Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	
gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tac atg	240
Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met	
65 70 75 80	
aga ggt gtt tgg ata gat act ggt gat aca cca aca gta ttt aga tat	288
Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr	
85 90 95	
att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta	336

Ile	Asp	Trp	Leu	Leu	Thr	Val	Pro	Leu	Gln	Met	Val	Glu	Phe	Tyr	Leu		
			100					105					110				
att	ctt	gcc	gct	tgt	aca	agt	gtt	gct	gct	tca	tta	ttt	aag	aag	ctt	384	
Ile	Leu	Ala	Ala	Cys	Thr	Ser	Val	Ala	Ala	Ser	Leu	Phe	Lys	Lys	Leu		
		115					120					125					
cta	gct	ggt	tca	ttg	gta	atg	tta	ggt	gct	gga	tct	gca	ggc	gaa	gct	432	
Leu	Ala	Gly	Ser	Leu	Val	Met	Leu	Gly	Ala	Gly	Ser	Ala	Gly	Glu	Ala		
		130				135					140						
gga	tta	gct	cct	gta	tta	cct	gct	ttc	att	att	ggt	atg	gct	gga	tgg	480	
Gly	Leu	Ala	Pro	Val	Leu	Pro	Ala	Phe	Ile	Ile	Gly	Met	Ala	Gly	Trp		
145					150				155						160		
tta	tac	atg	att	tat	gag	cta	tat	atg	ggt	gaa	ggt	aag	gct	gct	gta	528	
Leu	Tyr	Met	Ile	Tyr	Glu	Leu	Tyr	Met	Gly	Glu	Gly	Lys	Ala	Ala	Val		
				165					170						175		
agt	act	gca	agt	cct	gct	gtt	aac	tct	gca	tac	aac	gca	atg	atg	atg	576	
Ser	Thr	Ala	Ser	Pro	Ala	Val	Asn	Ser	Ala	Tyr	Asn	Ala	Met	Met	Met		
			180					185					190				
att	att	gtt	gtt	gga	tgg	gca	att	tat	cct	gct	gga	tat	gct	gct	ggt	624	
Ile	Ile	Val	Val	Gly	Trp	Ala	Ile	Tyr	Pro	Ala	Gly	Tyr	Ala	Ala	Gly		
		195					200					205					
tac	cta	atg	ggt	ggc	gaa	ggt	gta	tac	gct	tca	aac	tta	aac	ctc	ata	672	
Tyr	Leu	Met	Gly	Gly	Glu	Gly	Val	Tyr	Ala	Ser	Asn	Leu	Asn	Leu	Ile		
		210				215					220						
tat	aac	ctt	gct	gac	ttt	gtt	aac	aag	att	cta	ttt	ggt	ttg	atc	att	720	
Tyr	Asn	Leu	Ala	Asp	Phe	Val	Asn	Lys	Ile	Leu	Phe	Gly	Leu	Ile	Ile		
225					230					235					240		
tgg	aat	gtt	gct	gtt	aaa	gaa	tct	tct	aat	gct						753	
Trp	Asn	Val	Ala	Val	Lys	Glu	Ser	Ser	Asn	Ala							
				245					250								

<210> 55

<211> 251

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 55

Met	Gly	Lys	Leu	Leu	Leu	Ile	Leu	Gly	Ser	Ala	Ile	Ala	Leu	Pro	Ser
1				5				10					15		

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val

20	25	30
Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe		
35	40	45
Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr		
50	55	60
Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met		
65	70	75
Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr		
85	90	95
Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu		
100	105	110
Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu		
115	120	125
Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Ser Ala Gly Glu Ala		
130	135	140
Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp		
145	150	155
Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val		
165	170	175
Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met		
180	185	190
Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly		
195	200	205
Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile		
210	215	220
Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile		

225

230

235

240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 56

<211> 753

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone PalB6; GenBank # AF349998

<400> 56

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15

ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
 Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30

tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gtg ttc ttt 144
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
 35 40 45

ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act 192
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg 240
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80

aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat 288
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95

att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
 100 105 110

att ctt gct gct tgt aca aat gtt gct gct tca tta ttt aag aag ctt 384
 Ile Leu Ala Ala Cys Thr Asn Val Ala Ala Ser Leu Phe Lys Lys Leu
 115 120 125

cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct 432
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
 130 135 140

gga tta gct cct gta tgg cct gct ttc att att ggt atg gct gga tgg 480
 Gly Leu Ala Pro Val Trp Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
 145 150 155 160

tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta 528
 Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
 165 170 175

agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg gtg 576
 Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Val
 180 185 190

att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt 624
 Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
 195 200 205

tac cta atg ggt ggc gaa ggt gta tac gct tca aac cta aac ctt ata 672
 Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
 210 215 220

tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att 720
 Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
 225 230 235 240

tgg aat gtt gct gtt aaa gaa tct tct aat gct 753
 Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 57

<211> 251

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 57

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Cys Thr Asn Val Ala Ala Ser Leu Phe Lys Lys Leu
115 120 125

Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
130 135 140

Gly Leu Ala Pro Val Trp Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
145 150 155 160

Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
165 170 175

Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Val
180 185 190

Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
195 200 205

Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
210 215 220

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 58
 <211> 753
 <212> DNA
 <213> Naturally occurring gamma proteobacteria

 <220>
 <221> CDS
 <222> (1)..(753)
 <223> Proteorhodopsin variant from pcr clone PalB8; GenBank #AF350000

<400> 58
 atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15

 ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
 Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30

 tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gtg ttc ttt 144
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
 35 40 45

 ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act 192
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

 gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg 240
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80

 aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat 288
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95

 att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta 336
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
 100 105 110

 att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt 384
 Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
 115 120 125

 cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct 432
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
 130 135 140

 gga tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg 480
 Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
 145 150 155 160

tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta 528
 Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
 165 170 175
 agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg atg 576
 Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met
 180 185 190
 att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt 624
 Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
 195 200 205
 tac cta atg ggt ggc gaa ggt gta tac gct tca aac tta aac ctt ata 672
 Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
 210 215 220
 tat aac ctt gct gac ctt gtt aac aag att cta ttt ggt ttg atc att 720
 Tyr Asn Leu Ala Asp Leu Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
 225 230 235 240
 tgg aat gtt gct gtt aaa gaa tct tct aat gct 753
 Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 59

<211> 251

<212> PRT

<213> Naturally occurring gamma proteobacteria

<400> 59

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
100 105 110

Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
115 120 125

Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
130 135 140

Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
145 150 155 160

Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
165 170 175

Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met
180 185 190

Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
195 200 205

Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
210 215 220

Tyr Asn Leu Ala Asp Leu Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 60

<211> 753

<212> DNA

<213> Naturally occurring gamma proteobacteria

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone PalE1;GenBank# AF350001

<400> 60

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca	48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser	
1 5 10 15	
ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt	96
Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val	
20 25 30	
tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gtg ttc ttt	144
Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe	
35 40 45	
ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act	192
Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	
gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg	240
Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met	
65 70 75 80	
aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat	288
Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr	
85 90 95	
att gat tgg tta tta act gtt cca tta caa gtg gtt gag ttc tat cta	336
Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Val Val Glu Phe Tyr Leu	
100 105 110	
att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt	384
Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu	
115 120 125	
cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct	432
Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala	
130 135 140	
gga tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg	480
Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp	
145 150 155 160	
tta tac atg att tat gag cta tat atg ggt gaa ggc aag gct gct gta	528
Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val	
165 170 175	
agt act gca agt cct gct gtt aac cct gca tac aac gca atg atg atg	576
Ser Thr Ala Ser Pro Ala Val Asn Pro Ala Tyr Asn Ala Met Met Met	
180 185 190	

att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt 624
 Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
 195 200 205

tac cta atg ggt ggc gaa ggt gta tac gct tca aac tta aac ctt ata 672
 Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
 210 215 220

tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att 720
 Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
 225 230 235 240

tgg aat gtt gct gtt aaa gaa tct tct aat gct 753
 Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 61
 <211> 251
 <212> PRT
 <213> Naturally occurring gamma proteobacteria
 <400> 61

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Val Val Glu Phe Tyr Leu
 100 105 110

Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
 115 120 125

Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
 130 135 140

Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
 145 150 155 160

Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
 165 170 175

Ser Thr Ala Ser Pro Ala Val Asn Pro Ala Tyr Asn Ala Met Met Met
 180 185 190

Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
 195 200 205

Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
 210 215 220

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
 225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 62

<211> 753

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone PalE6; GenBank#AF350002

<400> 62

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15

48

ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt	96
Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val	
20 25 30	
tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gta ttc ttt	144
Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe	
35 40 45	
ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act	192
Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	
gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tac atg	240
Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met	
65 70 75 80	
aga ggt gtt tgg ata gat act ggt gat aca cca aca gta ttt aga tat	288
Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr	
85 90 95	
att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta	336
Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu	
100 105 110	
att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt	384
Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu	
115 120 125	
cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct	432
Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala	
130 135 140	
ggt tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg	480
Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp	
145 150 155 160	
tta tac atg att tat gag cta cat atg ggt gaa ggt aag gct gct gta	528
Leu Tyr Met Ile Tyr Glu Leu His Met Gly Glu Gly Lys Ala Ala Val	
165 170 175	
agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg aag	576
Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Lys	
180 185 190	
att att gtt att gga tgg gca att tat cct gct gga tat gct gct ggt	624
Ile Ile Val Ile Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly	
195 200 205	
tac cta atg agt ggt gac ggt gta tac gct tca aac tta aac ctt ata	672
Tyr Leu Met Ser Gly Asp Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile	
210 215 220	

tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att 720
 Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
 225 230 235 240

tgg aat gtt gct gtt aaa gaa tct tct aat gct 753
 Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
 245 250

<210> 63

<211> 251

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 63

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
 100 105 110

Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
 115 120 125

Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
 130 135 140

Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
145 150 155 160

Leu Tyr Met Ile Tyr Glu Leu His Met Gly Glu Gly Lys Ala Ala Val
165 170 175

Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Lys
180 185 190

Ile Ile Val Ile Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
195 200 205

Tyr Leu Met Ser Gly Asp Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
210 215 220

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250

<210> 64

<211> 753

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone PalE7; GenBank# AF350003

<400> 64

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
1 5 10 15

ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
20 25 30

tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gtg ttc ttt 144
Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
35 40 45

ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act	192
Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr	
50 55 60	
gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg	240
Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met	
65 70 75 80	
aga ggt gtt tgg ata gat act ggt gat acc cca aca gta ttc aga tat	288
Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr	
85 90 95	
att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta	336
Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu	
100 105 110	
att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt	384
Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu	
115 120 125	
cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct	432
Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala	
130 135 140	
gga tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg	480
Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp	
145 150 155 160	
cta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta	528
Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val	
165 170 175	
agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg atg	576
Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met	
180 185 190	
att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt	624
Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly	
195 200 205	
tac cta atg ggt ggc gaa ggc gta tac gct tca aac tta aac ctt ata	672
Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile	
210 215 220	
tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att	720
Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile	
225 230 235 240	
tgg aat gtt gct gtt aaa gaa tct tct aat gct	753
Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala	
245 250	

<210> 65
 <211> 251
 <212> PRT
 <213> Naturally occurring gamma proteobacterium

<400> 65

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
 1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
 20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
 35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
 50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
 65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
 85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
 100 105 110

Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
 115 120 125

Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala
 130 135 140

Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp
 145 150 155 160

Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val
 165 170 175

Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met
180 185 190

Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
195 200 205

Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
210 215 220

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245 250